ARL-TR-7715 ● JULY 2016

US Army Research Laboratory

Gesture-Based Controls for Robots: Overview and Implications for Use by Soldiers

by Linda R Elliott, Susan G Hill, and Michael Barnes

NOTICES

Disclaimers

ARL-TR-7715 ● JULY 2016

US Army Research Laboratory

Gesture-Based Controls for Robots: Overview and Implications for Use by Soldiers

by Linda R Elliott, Susan G Hill, and Michael Barnes
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
July 2016	Final	September 2013–July 2015

4. TITLE AND SUBTITLE
Gesture-Based Controls for Robots: Overview and Implications for Use by Soldiers

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Linda R Elliott, Susan G Hill, and Michael Barnes

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-HRS
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7715

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report provides an overview of gestural controls of robots in general, followed by a discussion of issues more specific to control of military ground robots by dismounted Soldiers. For the field of gestural controls, the technological progress is rapid and distributed among many different approaches, and the number of relevant publications is huge. A review of literature is provided, focused on 2 types of technological approach: camera-based and wearable instrumented devices. Handheld devices are also discussed in terms of augmenting gesture precision (i.e., pointing gestures). Attention is given to issues related to relative advantages of each approach for effective recognition and parsing of gestures, particularly in terms of their relevance to dismounted Soldier systems. Human-factors issues regarding the interaction of Soldiers and technology and effective design of user interfaces and controls are fundamental to successful use. This report identifies the major issues regarding applications to dismounted military operations.

15. SUBJECT TERMS
gesture, robot control, haptic, Soldier, Army

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Linda R Elliott
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	UU	70	706-315-1386

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

This report was accomplished through the support of the US Army Research Laboratory Human Research and Engineering Directorate programs regarding multiyear investigations of human-robot interaction.

1. Introduction

1.1 Background

A future vision of the use of autonomous and intelligent robots in dismounted military operations is for Soldiers to interact with robots as teammates, much like Soldiers interact with other Soldiers (Brown 2011; Lilley 2013; Phillips et al. 2013; Redden et al. 2013). Soldiers will no longer be operators in full control of every movement, as the autonomous intelligent systems will have the capability to act without continual human input. However, Soldiers will need to use the information available from or provided by the robot. One of the critical needs to achieve this vision is the ability of Soldiers and robots to communicate with each other. This report examines one mode of communication—gesture.

Part of the future vision includes bidirectional communication, with Soldiers and robots communicating with each other. However, this review focuses on human gestures to instruct and command robots. Therefore, we are describing (for the most part) one-way communications from human to robot. While this is very important, it is only one part of the larger vision for humans and intelligent, autonomous systems to interact with each other. Many efforts are focused on the use of gestures for robot control; in this report, we discuss technology options and issues impacting effectiveness for robot control in military operations.

The use of gestures as a natural means of interacting with devices is a very broad concept that includes a range of body movements, including movements of the hands, arms, and legs, facial expressions, eye movements, head movements, and/or 2-dimensional (2-D) swiping gestures against flat surfaces such as touch screens (Saffer 2008). Gesture-based technology is already in place and commonly used (e.g., public buildings, public restrooms) without special instruction required for effective use. A common example of a well-designed gestural command is the use of hands to "wave" to activate (e.g., public bathroom faucet). This concept is also common to gaming interfaces (e.g., Kinect) and is now extending to other private and public domains such as automobile consoles (Stokloso 2015).

The most cursory examination of the gestural literature shows great breadth and depth, with investigations of gestures arising from a variety of fields. Classification and interpretation of gestures have been discussed and reviewed over many years (Pavlovic et al. 1997; Moore et al. 1999; Kendon 2004). Gestures using hand motion are most common, and one useful classification scheme is by purpose: a) conversational, b) communicative, c) manipulative, and d) controlling (Wu and Huang 1999). Conversational gestures are those used to enhance verbal

communications, while communicative gestures, such as sign language, comprise the language itself. Manipulative gestures can be used for remote manipulation of devices or in virtual reality settings to interact with virtual objects. Controlling gestures can be used in both virtual and reality settings, and are distinguished in that they direct objects through gestures such as static arm/hand postures and/or dynamic movements.

Of particular interest in this report are applications and advancements with regard to controlling gestures for human-robot interactions. There have been many studies showing usefulness of more naturalistic interfaces for robot control (Goodrich and Schultz 2007). The study of gestures for robot control in itself is a huge field of endeavor. Gestures can be as simple as a static hand posture or may involve coordinating movements of the entire body (Yang et al. 2006). Gesture-based commands to robots have been used in a variety of different settings, such as assisting users with special needs (Jung et al. 2010), assisting in grocery stores (Corradini and Gross 2000), and home assistance (Muto et al. 2009). Examples of gestural commands in these settings include "follow me", "go there", or "hand me that". There are also advancements in various industrial settings to control robotic assembly and maneuver tasks (Lambrecht et al. 2011; Barattini et al. 2012).

Some aspects of gesture control are not included in this report. Stroke gestures made upon a screen (e.g., tablet, smartphone) represent a different domain of gesture control, which is also of interest to military human-robot interactions (O'Brien et al. 2009). Research regarding these stroke gestures attends to issues that define, develop, and validate approaches and taxonomies relating to the stroke gesture. This report does not address these issues, but rather is focused on free-form gestures made by the hand and arm, technology approaches to recognition of these, and how they may impact effectiveness within a military human-robot application. There is also interesting work to develop gestures for the robot to use to communicate back to the operator, such as head nodding (Muto et al. 2009; St Clair et al. 2011b), conversational gestures (Bremner et al. 2009), and queries (Iba et al. 2003). While bidirectional human-robot interaction is also pertinent to military settings, this realm of research deserves a separate review and is not addressed here.

1.2 Purpose

For this report, we start with a very general nontechnical overview of gestural controls of robots in general, then we narrow our focus to efforts more specific to control of military ground robots by dismounted Soldiers. We start with findings that are more general because certain features of gestural commands to robots are relevant to all settings: they should be intuitive, familiar, and easily distinguished.

Findings with regard to these different user groups can still generalize to military use, given the cross-cultural intuitive nature of some gestures (e.g., pointing). That is, many gestures can be intuitively recognized and used across population groups. Technical challenges for gesture-based control are also similar, regardless of operational setting. We review current progress and issues with a view of assessing different technological approaches for their relevance to dismounted Soldier systems. The technology approaches are very different and can strongly moderate effectiveness for different situations.

We cover 2 very different approaches to gestural control: camera-based systems and wearable instrumented systems. For the field of gestural controls, the technological progress is rapid and distributed among many different approaches within each general domain. In this review, we have attempted to include papers that represent a cross section of relevant approaches, by universities, government, industry, and countries, of varying disciplines and points of view. Our aim is to identify the major approaches and corresponding issues across the diverse range of gesture control endeavors. From this, we discuss characteristics of different gesture-based systems, with regard to capabilities, advantages, and limitations—as they pertain to use by dismounted Soldiers.

1.3 Gestural Control Task Demands

A front-end issue when considering use of any new system is consideration of the task demands and situational constraints. Here, we identify 5 types of tasks that can impact the choice of technological approach: simple commands, complex commands, pointing commands, remote manipulation, and robot-user dialogue. The developer should always start with a deep understanding of operational and situational task demands. In the following subsections we describe some basic tasks that distinguish the kind of technology and approach that will best meet the user requirements.

1.3.1 Simple Commands (Small Set of Specific Commands/Alerts)

While it is easy to think of single commands (e.g., "stop", "move forward", "turn left") as simple commands, one should keep in mind it is not the command per se, but the distinguishability and the intuitive nature of the gesture that determines ease of use and recognition. When the gesture set is small, recognition rates have been high, across many different camera and glove-based approaches. Gestures must be distinguishable from one another and, in general, gestures using movements of the arms as well as fingers have been more effectively recognized.

1.3.2 Complex Commands

Complex commands are characterized by higher demands for deliberative cognitive processing, often through use of a larger gesture set and/or combinations of gestural units to communicate multiple concepts. Any particular gesture set lies on a continuum of complexity (e.g., recognition difficulty), ranging from a small set of elemental gestures (e.g., "move forward", "turn left", "turn right", "stop") to gestures sets that may include large numbers of gestures, associated not only with meaning (i.e., vocabulary), but also elements of syntax and grammar, depending on sequence (e.g., American Sign Language [ASL]).

As the gesture set becomes larger, the challenge to effective recognition also increases. It is critical to have gestures that are distinguishable from the point of view of the recognition process; however, it is also desirable to have gestures that are intuitive to the command and easily learned (e.g., pointing). There is no distinct categorization of what comprises a small versus large gesture set, for the difficulty of recognition is also a function of distinguishability. It should simply be noted that these distinctions exist on a continuum, and this continuum of gesture set complexity will greatly moderate the results with regard to recognition effectiveness. This issue is very much related to concepts of visual salience (St Clair et al. 2011b) and tactile salience (Mortimer et al. 2011).

Scenario complexity is also increased through an increase in scenario actors (human and robot). Given a multi-entity situation, the goal is to attain robot awareness of the environment and actors, as interpreted through shared goals and perspective taking. This level of situation assessment and understanding by the robot is sought through development and application of agent-based and other computational decision models or cognitive architectures (St Clair et al. 2011a).

1.3.3 Pointing Commands

A fundamental task for gesture commands is that of directing movement for ground robots. Pointing gestures have been developed over several years, either to convey direction information or to clarify ambiguous speech-based commands. Perzanowski distinguished natural gestures (e.g., pointing a finger or whole arm) from "synthetic" gestures (e.g., use of a mechanical device such as mouse or stylus). The act of pointing is considered universally intuitive, as exemplified by attempts by infants to use pointing when trying to grasp objects out of their reach (Perzanowski et al. 2000b).

While the pointing gesture is natural and intuitive, recognition of "where" and "what" can be challenging, depending on task context. When using a pointing gesture, there should be a distinction between pointing to a specific object (e.g.,

"please bring me that cellphone"), pointing to a specific location (e.g., "through that alley") and pointing to a generalized area (e.g., "search that area"). Areas can be more precisely circumscribed when augmented by use of a map display (e.g., circling the area of interest) (Brooks and Breazeal 2006; Perzanowski et al. 2000a, 2000b). Other approaches have used object recognition as an aid to gesture interpretation (e.g., "bring me that cup"). A different approach was taken by Strobel et al. (2002), where they used the spatial context of the environment (e.g., orientation of object to axis of hand) to help disambiguate hand gestures. Advancements in instrumented glove technology are enabling determination of azimuth from a point gesture; when combined with GPS-based wearable device, both direction and distance can be determined through sensors within the glove (Vice 2015).

Littmann et al. (1996) discuss progress on a point and place task for a robot capable of grasping maneuvers (e.g., "pick up that object and place it over there") using 2 color cameras that provide stereo data. The output is further processed by color-sensitive neural networks (NNs) to determine object location. In a laboratory setting, the accuracy of the system for localization was that of 1 cm in a workspace area of 50 × 50 cm. The 2 cameras must show both the human hand and the table with objects. In addition, each training exemplar must show a hand-pointing posture associated with a known location.

Abidi et al. (2013) used a Kinect sensor to interpret pointing gestures that directed a robot to move to a location that the user pointed to on the ground. The 3-D location of a person's right elbow, right hand, and eye gaze were captured to determine location. Two approaches were compared: gesture alone and gesture combined with eye gaze. While participants expressed a strong preference (i.e., 62% vs. 38%) for the combination of gesture with eye gaze, no performance data were reported with regard to accuracy of localization by the robot. An advantage to this approach is that the robot continually assesses the pointing gesture while moving, allowing for real-time change; however, the robot must always keep the operator in sight. This prevents pointing to a location that is outside this range.

1.3.4 Remote Manipulation

Ground-based mobile robots are often used for remote manipulation of objects. In combat situations, this capability is often used for bomb disposal (Axe 2008). Several efforts have been reported where gestures have been developed for remote manipulation. Several of these regard the development of service robots designed to assist people in locations such as offices, supermarkets, hospitals, and households. Other efforts focus on assisting users in more dangerous environments

such as hazardous areas or space, using telepresence and teleoperation (see Basanez and Suarez [2009] for a review of teleoperation issues).

One primary manipulation common to most applications is that of grasping. Grasping consists of several steps: a) perception of object, b) determination of object form, size, orientation, and position, c) planning the grip, d) grasping the object, e) moving the object to a new location, and f) releasing the object (Becker et al. 1999). Becker and his associates developed a camera-based system that allows interpretation of gestures that indicate the choice of object, the grip to be used, and the desired final position. The task environment was restricted to that of a tabletop. The camera was a dual-stereo camera with 3 degrees of freedom (DOF) (pan, tilt, and vergence) with both color and monochrome capabilities. Recognition software was based on the C++ library FLAVOR ("Flexible Library for Active Vision and Object Recognition") developed at the Institut fur Neuroinformatik (Bochum, Germany). Tracking of the operator's hand relied on motion detection, skin color analysis, and a stereo cue. Object recognition was based on object edges.

Colasanto et al. (2013) describe issues and challenges inherent in developing successful teleoperation of anthropomorphic robotic hand-arm systems. They support the approach where the motion is made with the human hand and captured, to enable proper imitation by the robotic hand. While visual-based systems have been used for grasping tasks, they usually require special markers on the hand and still have problems due to visual occlusion. They describe algorithmic approaches that used an instrumented wired glove (i.e., Cyberglove having 22 joint-angle measurements) and a 4-finger robotic hand. They describe development of an effective approach combining joint-space mapping and fuzzy-based pose mapping that accommodated both grasping movements and more free-form movements.

Stiefelhagen and associates (2004) reported the development of a system using speech, gestures, and head pose to accomplish grasping tasks such as retrieving objects, turning a lamp on/off, or setting a table. While they report correct interpretation of each component (e.g., speech, head orientation, pointing gesture) and of the combination of components, they did not report performance data on the manipulation tasks per se. Lii and his associates (2012) describe uses of a library of tasks linked to gesture commands given by an operator wearing a Cyberglove. They describe teleoperation of grasps and manipulation with a 15DOF robot hand and 6DOF object manipulation, using different grasp approaches.

Camera-based approaches have also been used for remote manipulation. Lathuiliere and Herve (2000) reported successful real-time application using a single camera, where the operator wore a dark glove marked with colored cues. This approach was used for 2 different grasping maneuvers (e.g., spherical and cylindrical). Raheja et

al. (2010) also used a camera-based approach for robotic hand control. Using a technique described by Raheja et al. (2009), to manage light illumination and viewing angle, the researchers developed a system to recognize hand gestures. Eleven different gestures were used to control a robotic hand. The gestures were restricted to static posturing of the hand (e.g., clenching a fist, spreading of fingers, thumbs up/down, 2-finger "peace" sign, 1 finger, 3 fingers). Recognition was reported at 90% but was very dependent on proper light illumination. Celik and Kuntalp (2012) also used a camera-based approach for control of grip commands and compared 2 methods for gesture recognition. The template-matching algorithm, based on pre-stored gestural information, was found to be faster than the signature signal algorithm (based on identification of edge data) but had higher memory requirements. Their results appear to be based on 2 commands (open gripper, close gripper), therefore these differences may compound as the number of gestures increase.

Researchers from Keio University in Yokohama, Japan, investigated user preferences for gestural control of a robotic "hand" for assistive tasks such as assembly (Wongphati et al. 2012). Control movements included 6 direction movements and open-close gripper movements that were applied to assist users with a soldering task. While many gestures are developed from the point of view of engineering (e.g., ease of detection), in this study the focus was on capturing gestures that were more intuitive to the user. While results are preliminary and exploratory, trends indicate user preference to use one or both hands, as opposed to other body parts/movements.

An interesting application reported by Park et al. (2005b) is that of a "competition" robot that could make fighting maneuvers (stand up, hook, turn left, turn right, walk forward, back pedal, side attack, move to left, move to right, back up, both punch, left punch, right punch). Pose extraction was accomplished through skin color regions using a 2-D Gaussian model that extracted positions of the face, left hand, and right hand. Hidden Markov modeling (HMM) was used to process the continuous stream. Each gesture was performed 100 times by 10 different individuals. Reliability of recognition was reported to be 98.95%.

1.3.5 Robot-User Dialog

As robots become more autonomous, command of the robot transitions from direct and detailed teleoperation commands to higher-level command dialogs. Several approaches have been taken to establish dialogs that allow robots to acknowledge commands and ask for clarification. At the least, robots have to inform users of their status, through visual, auditory, or other means. International standards exist for visual signals of danger and for auditory alarms. However, there is increasing

need for more complex dialog to attain clarity (e.g., "you told me to go somewhere but you did not say where"), (Kennedy et al. 2007; Perzanowski et al 2000a, 2000b). Many efforts are currently focused on developing the ability for the robot to not only maintain a dialog, but to learn new interactions over time, such as use of gaze direction or pointing gestures toward an object of interest (Ou and Grupen, 2009). Taylor and associates report progress for greater dialog based on a computer programming architecture (i.e., Soar) for artificial intelligence (Taylor et al. 2012). Soar provides a robust framework for knowledge representation and logic (i.e., "if-then" rules). These efforts represent growing focus and progress with regard to communications based on logic and reason.

In addition to enhanced clarity and complexity of rational commands, there are also many efforts focused on social aspects of human-robot communications. Muto et al. (2009) explored options for human-robot dialogue for situations involving a social robot to assist elderly users. Findings from an investigation of human-human dialogue were applied to develop and refine the robot's response to a user request, such as "could you give me a wood block". Human-human interactions were observed, where the request utterance speed was varied. The average pause between request and response was 300 ms. In subsequent human-robot interactions, speed of the request was varied, along with the robot's response, using speech and gesture (i.e., head nod, grasping), with regard to timing, order, and length of pause after the request. Afterward, these interactions were rated by the human participant. They found that older participants favored the condition where the robot response timing matched the timing of the request utterance (e.g., fast, slow). This preference was not demonstrated by younger participants. The use of the older versus younger participants was exploratory; there were no a priori hypotheses with regard to expected differences. However, researchers suggested differences might be in part due to differences in subjective time perception. It may also be due to younger participants having higher levels of familiarity with technology and robots. Findings suggest the need for more research in this area.

While much of the progress in human-robot dialog is based on speech, there are complementary efforts to incorporate gestures to enhance, clarify, and/or replace speech, when appropriate. St Clair and his colleagues (2011b) investigated human-robot effectiveness in visually cluttered environments to better understand the effect of visual saliency on gesture production by a robot. In this case the gestures of interest were those indicating a distal location or item to elicit clarification from the operator. Pointing gestures produced by the robot included orientation of the head (e.g., "eye" gaze direction), with straight arms, with bent arms, and the combination of arm and head gestures. Target items were varied with regard to visual salience, distance, angle, and pointing modality. Results showed that the

pointing modality affected results to a greater extent than that of visual salience. The combination of head orientation and arm gesture was more accurate than either gesture was alone.

Taylor and his associates (2014), in their development of gestures for a robotic mule, offer a useful taxonomy for the classification of gestures types, which includes the following:

- Static versus dynamic: Whether the gesture is a "pose" or requires movement.
- Continuous versus discrete: Whether the gesture can be repeated and still be recognized.
- 1-arm versus 2-arm: Some are made with one arm, others with two.
- Inclusion of hand pose: Whether the gesture is arm-only or also includes a hand position.
- Inclusion of body reference: Whether the gesture is hand-arm only, or refers to another part of the body (e.g., tapping the head).
- Facing toward or away from target of communication: Some gesture recognition systems require facing toward the receiver.
- Gestures in x-y plane versus x-y-z plane: Some gestures may have more complex movement in 3-D space.
- Unique versus dependent: Whether gesture has only one meaning or is dependent on context.

These distinctions become more useful and relevant as other issues are also considered (e.g., purpose of gesture, type of recognition system).

1.3.6 Summary

Designers and managers must give careful consideration to operational task demands to develop or select the appropriate system for diverse commands, ranging from simple commands to more complex human-robot dialogs. The need for remote manipulation or accurate communication of azimuth and distance (i.e., pointing) introduces additional challenges and considerations. Both camera-based and instrumented systems vary in effectiveness for the different types of task demands. All varieties have demonstrated effectiveness for simple commands. All gesture systems are more challenged when accomplishing pointing gestures, and may need augments such as visual display, speech, or use of a handheld pointer, to better localize the target. Instrumented gloves can have an advantage over camera-based

systems, to accomplish more complex and/or manipulation commands, particularly in situations having high visual clutter or greater distances between the operator and robot. Future concepts include more bidirectional dialogs between the operator and robot. In addition, operational context, such as visibility, weather, and need for colocation or direct view of the robot, will also impact the effectiveness of a given gesture system.

2. Camera-Based Systems

2.1 Overview and Options

Camera-based gesture recognition systems have used a variety of camera types to capture images that are then interpreted with some form of quantitative algorithmic interpretation of the video (Hassanpour et al. 2008). This type of approach is also referred as vision-based recognition (Gavrilla 1999; Wu and Huang 1999). Applications have ranged from a small set of body postures to large sets of complex hand gestures (e.g., ASL). A well-known application is that of the Kinect system, where camera-based interpretation of user body posture and movements serves to control videogame features and feedback.

It should first be noted that camera-based systems vary widely—with regard to type of camera sensor, number of cameras, and various algorithms used for gesture recognition. More recent advancements are as follows.

2.1.1. 3-D Data Recognition

More-recent improvements to camera-based recognition systems strive to better capture 3-D movement. Progress in this area has been achieved over time. Early attempts were fairly bulky (Kortenkamp et al. 1996). The 3-D camera sensors offer an alternative that may alleviate some of the problems with 2-D camera displays. Vogler and Metaxas (2001) developed parallel HMMs to recognize ASL gestures using a 3-D camera system. Xie et al. (2010) describe a stereo vision-based approach to identify the trajectory of a dynamic gesture, reporting a recognition rate of 92%. Gordon et al. (2008) used 2 cameras to achieve stereo vision to attain depth information. The 3-D approach to hand postures needs a large enough data set to cover the large number of possible hand shapes from various viewing angles. While successes have been reported with regard to 3-D-based real-time recognition of human shapes and movements, results are limited, restricted, and more relevant to surveillance as opposed to robot control per se (Checka and Demirdjian 2010; Cohen 2013).

The 3-D sensors with additional depth perception data (e.g., Kinec, etc.) offer even more detailed data for interpretation. Use of depth perception data allows the user more range of motion that can be interpreted by the system. Several studies have used the Kinect depth sensor system (Cheng et al. 2012). Yanik et al. (2012) used Kinect technology to build recognition of 3 hand signals taken from ASL to command assistive robots (i.e., “come closer”, “go away”, “stop”). Their approach used the Growing Neural Gas (GNG) algorithm that is more robust to variations in gesture execution. Skeletal depth data collected by the Microsoft Kinect sensor was clustered with the GNG. They report initial progress on their goal to develop a system that will improve with user feedback. Lai et al. (2012) reported 2 approaches to gesture recognition of 8 hand gestures, both using Kinect depth data and both resulting in over 99% accuracy of real-time recognition. While both approaches were accurate, the one based on Euclidean distance was associated with more limitations. Barattini et al. (2012) developed a gesture set for control of industrial robots, based on distinguishability data (i.e., confusion matrix), as determined based on use of the Microsoft Kinect system and dynamic time warping (DTW) recognition algorithms. Others have also applied the Kinect capabilities for industrial robots (Lambrecht et al. 2011) and for humanoid robots (Suay and Chernova 2011). Masum et al. (2012) applied Kinect capabilities for whole body gesture recognition, for the naturalistic control of a humanoid robot. They reported 99.87% accuracy using the Fuzzy Neural Generalized Learning Vector Quantization algorithm. Konda et al. (2012) used depth perception data to better adapt to outdoors conditions.

Xu et al. (2012) described the development of a gesture control system for robot control, using Kinect depth perception data for the control of a human service robot. HMMs, using left-right topology to identify hand trajectory, are used to model and classify 3-D tracking of 2 hands to distinguish 1) 6 navigation commands used by the right hand: turn right, turn left, move forward, move backward, rotate, and brake and 2) use of the left hand for velocity (robot speed) control. The system was able to recognize gestures by both hands at the same time. Start and end of a gesture was determined through a detection method combining an initial region with hand velocity. While assessment of real-time performance was accomplished in laboratory settings, the system was stated as robust to changes in illumination, color variations, and background clutter. The system was evaluated using gestures generated by 6 people that each performed 60 gestures. Gesture recognition rates ranged from 96% to 100% for the right-hand gesture, and 100% with the left hand. These rates were compared to recognition rates based on skin color segmentation (Elmezain et al. 2008) and DTW (Wu et al. 2010). While statistical significance is not reported, recognition rates were similar or higher using the depth data approach, and lower using the dynamic time warping approach. Recognition time was stated

to be based on time to complete a gesture, about 0.5 s. Navigation accuracy was described through diagrams representing ground truth (i.e., the path that the robot was supposed to traverse) and actual performance. Diagrams were quite similar in pattern; however, there was no performance data with regard to navigation errors.

Li and Pan (2012) used Kinect technology (e.g., skeleton tracing) and a DTW recognition algorithm for real-time control of a ground robot using 11 arm-based gestures (e.g., left hand to right; right hand to left, 2 hands zoom in, 2 hands clap, etc.). Using 5 people, with each person demonstrating each gesture 20 times, they reported accuracies from 90% to 96%. They stated response times were short but did not report actual times. They also reported that the user was able to use real-time returned video to manipulate the robot to avoid obstacles and successfully reach its destination.

Sugiyama and Miura (2013) reported an approach that integrates a head-mounted camera with a 9-axis orientation sensor and hand-worn inertial sensors. Sensor information is fused to estimate walking motion and hand motions, which then drives the movement (walking and arm motion) of a humanoid robot. The camera detects the hand through skin color and finger edges and can detect movement and grasping gestures. The user was able to successfully control robot movements. The primary application was that of remote telepresence control. Gopalakrishnan et al. (2005) discussed how camera-based systems could be augmented by integration with laser-based localization, a visual map display, and speech recognition capabilities.

2.1.2 Multicamera Networks

Multicamera networks show much promise with regard to recognition of a variety of types of gestures in a complex environment (Aghajan and Wu 2007); however, feasibility of such networked systems in operational environments is limited. One application that has reported success with this approach used multiple cameras to control a robotic forklift, such that the object can be recognized, and after a period of time, reacquired, even after the robot has moved long distances. First, the operator gives the robot a “guided tour” of named objects and the locations. Then the operator can dispatch the robot to fetch a particular object by name (Walter et al. 2010).

Park et al. (2005b) used a multicamera system to enable 2 robot controllers to participate in an indoor robot competition that involved fighting movements between 2 small humanoid-like robots (e.g., similar to transformer toys). This camera-based approach used recognition of skin color regions (i.e., face, left and right hands) of each operator for pose extraction and HMM-based gesture

recognition. Each of the 13 gestures was performed 100 times by 10 different operators, resulting in an overall reliability of 98.9%.

2.1.3 Recognition Algorithms

At this time, several reviews agree that the main technological challenge of camera-based systems rests with the efficacy of feature capture and gesture recognition processes, such as the review by Rautaray and Agrawal (2012). In another recent review, Khan and Ibraheem (2012) describe these phases as extraction method (e.g., image capture), features extraction, and classification (e.g., recognition). Image capture approaches include 2-D monocular systems, 3-D stereo systems, 3-D systems with more advanced depth information (e.g., Kinect), and multiple-camera systems. While capabilities among these systems vary, they each face similar challenges and assumptions, such as the need for high-contrast backgrounds under optimal ambient lighting conditions. An appearance-based approach uses a simpler 2-D model; still, extracting the image of hand movements in a cluttered environment can be challenging. Murthy and Jadon (2009) also provided a review of issues related to vision-based gesture recognition, noting that such systems are more effective in controlled environments. Problems arise from less-than-optimal illumination and visual noise (Kang et al. 2009).

Camera-based gesture recognition relies on some type of algorithmic approach to extract, parse, and classify gestures. The challenge to all approaches is gesture-spotting, which is the extracting of the start and stop points of specific gestures from a continuous stream of dynamic gestural movement (Alon et al. 2005). At this time, all approaches are associated with some level of error. Recognition of dynamic gestures has often been accomplished with some variation of HMM. HMM is a statistical procedure that builds upon Markov models, in that they can make inferences based on data that are probabilistic and sequential, such as speech recognition (Baker 1975). Many refinements have been explored and applied, finite state machines (FSMs), artificial neural networks (ANNs), continuous dynamic programming (Alon et al. 2005), and DTW (Kobayashi and Haruyama 1997; Shah and Jain 1997; Wu and Huang 1999; Corradini and Gross 2000; Urban et al. 2004; Li and Pan 2012). Rao and Mahanta (2006), instead of analyzing all frames of a video feed, applied methodology to extract and analyze a subset of key frames. They also used a clustering algorithm on static gestures and reported success on 5,000 gestures, with recognition rates from 84%–100%. Similarly, Chian et al. (2008) reported a method to more efficiently parse gesture classifications. Corradini and Gross (2000) compared different algorithms with regard to recognition rates for a set of 5 basic gestures. These included combination NNs with HMM, combination of radial basis function (RBF) with HMM, and recurrent

NN. While the HMM/RBF approach had somewhat higher recognition rates, the data were insufficient to draw conclusions other than that all approaches were associated with fairly high recognition rates associated with a small set of gestures, accomplished in laboratory conditions.

Most approaches to camera-based recognition rely on preprogrammed algorithms based on extensive repetition of a small set of gestures. However, there are attempts to make the recognition process more dynamic and user friendly. Hashiyama et al. (2006) used camera-based information in such a way as to allow users to create their own gestures for specific commands. Users "show" the gesture to the robot system, which then learns to recognize the gesture, for a particular command. Users were able to accomplish this within 30 min.

To summarize, many different recognition algorithms and strategies are currently being investigated. No one particular approach has been established as best; instead, it is likely that ideal solutions will be situational, depending on factors such as situation context (e.g., indoors, lighting), number of gestures, and type of gestures (e.g., range of movement). In the following section we discuss some factors contributing to recognition effectiveness.

2.2 Military Applications: Camera Systems

Camera-based gesture control systems have been investigated for several military applications, including robot control. Other applications are also included in this section, as issues regarding use and effectiveness generalize among the various task demands.

2.2.1 Dismount Soldier Communications

Cohen (2000, 2005) developed and demonstrated a prototype gesture recognition system for dismounted Soldiers based on camera-based gesture recognition of existing Army hand and arm signals (e.g., Army Field Manual 21–60 [Headquarters, Department of the Army 1987]). The approach for this technology was developed and demonstrated previously (Beach and Cohen 2001). The purpose was to prove the capability for technology-based recognition of Soldiers using a subset of Army hand and arm signals. The capability was developed to enhance training of gesture-based communications. The system would monitor if the proper gesture was performed adequately and if not, show how to perform the required gesture. The system also had a goal to be able to learn new gestures. Dynamic gestures included "slow down", "prepare to move", and "attention". Static gestures included "stop", "right/left turn", "okay", and "freeze". In this case, the dynamic gestures included the actual movement of the arms as part of the gesture. It is also

possible for camera-based systems to recognize static gestures, where a single, static position conveys the gesture meaning, and does not include the actual movement as part of the gesture. They also used a variety of other gestures, based on circles and lines, to check for recognition performance. Recognition rates varied from 80%–100%, with many gestures recognized at 95%–100% accuracy. This camera-based approach to recognition of gestures, motion tracking, and feature matching has been applied to numerous surveillance applications (Cohen 2013).

2.2.2 Robot Control

Perzanowski and his associates (Perzanowski et al. 1998; 2000a, 2000b; 2002, 2003) reported progress toward a multimodal approach to control of single and multiple robots using gesture, personal digital assistant (PDA), and speech. Syntactic and semantic information is drawn using ViaVoice speech recognition and natural language understanding system, Nautilus (Wauchope 1994). Visual cues include body location, eye gaze, or other types of body language. This is supplemented by recognition of gestures, such as pointing. In the 2002 instantiation, gesture recognition was based on a camera with a structured light rangefinder mounted to the side of the robot to track the user's hands, while sonar sensors are used to detect objects in the environment. In 2003, a Wizard of Oz (WoZ) experiment was conducted to explore naturally occurring preferences with regard to the use of gestures and language syntax. Natural language and spatial relationships are based on an approach described by Skubic (Skubic et al. 2001a, 2001b, 2002, 2004). Skubic and her research associates explored linguistic spatial relationships and directives (e.g., "go around the desk and through the doorway") when referring to an evidence grid map. The evidence grid map is built from range sensor data, where objects are assigned labels provided by the user. Robot software includes spatial reasoning that enables the robot to understand linguistic descriptions (e.g., "object is behind the desk") or commands (e.g., "go through the doorway then stop").

Sofge et al. (2003a) described an agent-based approach to control an autonomous robot using natural language, spatial reasoning, and gesture interpretation. The gestural system included a structured-light rangefinder that emitted a horizontal plane of laser light and a camera mounted on the robot with an optical filter tuned to the laser frequency. The camera generated a depth map from the reflection of the laser off objects in the room. Hand gestures were recognized because they were closer to the camera than other objects and were then processed to generate trajectories that are used for gesture recognition. Gestures indicated direction, and were integrated with speech commands such as "go over there". The focus of this effort, sponsored by Defense Advanced Research Projects Agency (DARPA), was

on the use of agent-based capabilities to integrate screen-based, visual, and gestural commands along with object recognition and spatial reasoning.

Brooks (2005) and Brooks and Breazeal (2006) reported progress toward naturalistic interaction with robots for Soldier tasks, which included robot capability to learn and imitate tasks, camera-based recognition of social interaction cues (e.g., gaze direction, nodding, facial expressions, etc.), and codified verbal expressions. In this case, the robot is not controlled directly by gestures; instead, the robot visual attention system attempts to monitor and recognize gestures and facial expressions of the operator to ascertain stimuli of interest. Robot capabilities included detection of head orientation, body motion mimicry, hand reflex, figure-ground segmentation, and response to operator gestures and touch. In addition, gesture recognition is integrated with a representational language for humanoid movement, with the goal of mimicry. Operational goals include scenarios where the user can show the robot what to do (e.g., open a gas tank, stack boxes, etc.).

Kennedy et al. (2007), using ViaVoice, programmed speech and gesture commands for a core set of robot control commands relevant to Marine reconnaissance missions: "attention", "stop", "assemble" (i.e., come here), "as you were" (i.e., continue), "report" (which assumes the robot can communicate to the user). The first 4 commands were taken from the US Marine Corps Rifle Squad manual (Headquarters, Department of the Navy 2002). The robot interacted with a team member through voice, gestures, and movement, using artificial intelligence programming (i.e., ACT-R) capability with additional capacity for spatial reasoning and perspective taking. Together, the team member and robot were tasked to covertly follow and approach a moving target (e.g., another human or robot). The target continually moved to various locations and had a limited field of view in which to spot any followers. The robot had logical reasoning to enable covert approach, such as "if the target is on the north, east, south, or west side of an object, it should hide on the opposite side of the object". While the emphasis of this effort was on spatial reasoning, gestures and speech were used for intercommunications. However, it is not clear from the report as to the extent to how many user-to-robot communications were used or successfully interpreted.

Jones (2007) reported development of a prototype robotic system capable of detecting and following a person through indoor and outdoor environments while responding to voice and gesture commands. A camera-based system was mounted to the arm of a Packbot, placing the camera about 5 ft above the ground, allowing the view of a person's upper body and gestures. The sequence of observed arm poses was matched to a complete sequence corresponding to a known gesture (e.g., wait, follow, breach doorway) through HMM algorithms (Rabiner 1989).

Ruttum and Parikh (2010) reported development of a gestural robot control system using core hand and arm signals used by the Marines. They focused on 4 signals, “come”, “go here/move up”, “stop”, and “freeze” and identified distinguishing factors from arm, hand, and body orientation/velocity/acceleration. Their approach was camera-based with real-time analysis of continuous video based on a Vicon motion detection system. This system is stated to reduce the limitations with regard to background clutter and orientation of the person to the camera. The assessment was conducted indoors (in an area measuring 30 × 25 × 9 ft). The subject was tagged with markers that are detected by infrared cameras set up within the room. They also compared 2 methods of analysis: Bayesian and NNs. Preliminary results were reported as inconclusive; however, the authors stated higher expectations for the NNs as data become more complex.

Advancements in camera-based interpretation of human movements have evolved to enable recognition and tracking (Gavrilla 1999). Kania and Del Rose (2005) describe successful application of camera-based techniques to detect pedestrian movements and thus augment robot leader-follower performance.

Camera-based recognition concepts have been demonstrated for leader-follower tasks, and other simple gestural commands, as shown in Fig. 1. There is also the potential use of more advanced applications of camera-based recognition, as it more closely approximates actual computer vision. For example, given surveillance as a mission goal, a camera-based computer vision system can serve to interpret surroundings in the environment, as well as the operator gestures (e.g., threat assessment based on motion, etc.) (Cohen 2005, 2013).

Fig. 1 Camera-based control of robotic mule (Taylor et al. 2012)

2.2.3 Aircraft Direction

There are ongoing efforts to build gesture recognition for aircraft and unmanned aerial vehicle (UAV) handling, which have similar task demands to ground robot control. Recent efforts to develop camera vision-based recognition of aircraft handling hand and arm signals have included work by Choi et al. (2008), as well as

a current Office of Naval Research–funded research and development effort being conducted at the Massachusetts Institute of Technology by Song et al. (2011a, 2011b, 2012). Choi et al. report overall accuracy of 99%, but for a training set consisting of multiple repetitions of only 8 gestures, while Song et al. have demonstrated gesture recognition accuracy of 75.37% for a subset of 24 aircraft handling gestures. In both cases, data were collected within a highly controlled laboratory setting in which lighting was controlled, visual noise was eliminated, an optimized field of view (FOV) and distance to the user were ensured, and hand and arm signals were generated by individuals that were standing still, rather than interacting within a complex and dynamic environment. Thus, it is unclear how well these systems would operate in challenging circumstances to meet operational needs.

Ablavsky (2004) also reported progress toward proof of concept with regard to the use of a camera-based gesture recognition system for the direction of UAV movements on aircraft carrier decks. The passive camera system used a wide FOV for recognition of blinking beacons and a narrow FOV for observing hand and arm gestures. In contrast, Urban et al. (2004) used a motion tracker along with wearable sensors toward the same task goals. Sensors were attached to armbands worn by the operator. Two sensors on each arm (i.e., upper and lower arm) were determined to be sufficient for most gestures in the Navy gesture lexicon for control of aircraft on the ground. While gestures were accurately recognized, there were problems associated with wired sensors (e.g., tangled wires, necessity of being near the base device) and fatigue associated with bulky sensors, indicating the need for smaller, wireless sensors.

2.2.4 Camera-Based Gestural Systems: Constraints in Military Operations

There are a number of things to think about when technology is considered for use in military operations. Military operations have a number of characteristics that will be more challenging than simple tasks in controlled settings. Camera-based systems that have been shown to be successful in controlled laboratory settings may not generalize to military operations. The following bullets discuss how camera-based gesture recognition systems might fare under some military operations circumstances.

- Line of sight. A primary constraint is line of sight: the camera and operator must be within sight of each other.
- Visual clutter in operational environments. Clutter represents increased challenges due to distance and angle from the camera, varied contrast

backgrounds, and visual degradation from smoke, exhaust, haze, and inclement weather.

- Visual clutter due to additional personnel. Camera-based systems may have difficulty recognizing the operator if other personnel (e.g., squad members) are within the visual field.
- Visibility. Camera-based recognition is degraded or ineffective in poor visibility due to smoke, fog, and night operations. This can be ameliorated with infrared or thermal cameras.
- Complex fast movements. High variation of body movements, the self-occlusion of one body part by another, and fast movements of arms and legs (Checka and Demirdjian 2010; Yanik et al. 2012).
- Latency. Latency can be a particular problem when operational tempo is fast and movements are quick.
- Generalizability of recognition to a wider range of users remains uncertain as the performance metrics are typically based on a training set developed from a small number of participants in ideal circumstances (Garg et al. 2009).
- Complex environments. The hand gesture detection methods that are based on skin color, 2-D, or 3-D template matching are not sufficiently robust due to the many degrees of freedom with regard to hand movements, unobvious exterior features of hands, illumination changes, and varied colored and cluttered backgrounds (Xu et al. 2012).
- Multiple cameras are not easily used on the go.

Table 1 lists these constraints, along with impact on operations and ways in which the constraint can be managed.

Table 1 Typical constraints in military operations associated with camera-based gestural systems

Issue	Operational impact	Potential solutions
Line of sight	Must use within line of sight (e.g., leader-follower scenario)	Future: multiple cameras/perspectives
Visual clutter	Use in noncluttered environments (e.g., roads, simple terrain)	Augment with GPS, other wearable markers
Visibility	Limit to high-visibility operations/daytime use	Augment with infrared/thermal cameras
Complex/fast movements	Limit to few simple commands/keep operator separate from others	Augment with GPS/wearable markers
Latency	Do not use in high-tempo operations	Technology improvements
Generalizability to other operators	Train camera/operator as a system	Algorithm improvements/wearable markers
Multiple cameras	Static situations (e.g., indoors)	Technology improvements for integration of multiple cameras on the go
Pointing gestures	Pointing not well accomplished	Augment with speech recognition; Improved technology

2.2.5 Approaches to Enhance Camera-Based Recognition

A strong advantage of camera-based systems in military operations is the direct link between the camera and the operator. No wireless transmissions are needed to achieve communications, thus avoiding issues regarding signal strength or jamming. This issue can be the deciding factor for some military situations. Potential advantages also arise when camera-based vision systems serve multiple purposes, based on general recognition, not only of gestures, but of objects (stationary and moving) and potential threats.

However, camera-based system effectiveness can be degraded in cluttered environments or if the operator is out of the line of sight. There are a number of approaches used to enhance camera-based recognition. As discussed in the following paragraphs, these approaches include making the gesture more "visible" (e.g., wearable markers), controlling the environment, or enhancing the camera technology.

2.2.5.1 Wearable Markers

A major challenge in more complex settings is the capability of recognizing a discrete gesture from a continuous stream of motion against a cluttered background. Some success has been reported using real-time continuous data (Alon et al. 2005).

Some progress has been made with regard to vision-based gestures in loud and cluttered settings, where user-borne devices such as microphones or gloves were not an option (Barattini et al. 2012).

The efficacy of camera-based gesture recognition can be aided by use of markers (e.g., special clothing, colors, focus on skin tones, etc.) to simplify and speed the recognition process. Waldherr et al. (2000) used face color and shirt color as features to track movements. Manigandan and Jackin (2010) used skin color and textures to simplify recognition of hand postures. Singh et al. (2005) had the user wear clothes with colors that stand out from the background, with recognition based on motion and color cues and accuracy up to 90% for 11 commands. Malima et al. (2006) reported fast and efficient recognition of "counting" gestures through segmentation of the hand based on skin color and size. Stancil et al. (2012) described development of a robotic mine dog (Neya Systems LLC) that located its operator through recognition of body shape, posture, and a jacket worn by the operator.

2.2.5.2 Controlled Environment

Demonstrations of camera-based recognition often occur in controlled laboratory settings. Recognition can be aided through minimization of movement of all other objects in the field (Singh et al. 2005).

2.2.5.3 Adding Speech Recognition

Speech recognition has been used to facilitate gesture systems of all types. Several Department of Defense efforts are focused on integration of gestures with speech for robot control. While there are advantages to the independent use of gesture-based control, there are also advantages that have been demonstrated when gestures are integrated with speech to achieve communication that is both naturalistic and precise. Speech is a very natural means of communication and is often the preferred modality of control (Beer et al. 2012). Progress toward natural speech control has been demonstrated for situations involving autonomous ground robots (Schermerhorn 2011; Hill et al. 2015).

Jones (2007) combined camera-based recognition based on analysis of 3-D point clouds with speech recognition. Robot capabilities include that of person-following, gesture recognition based on silhouette and 3-D head position ("wait-follow", "breach doorway"), and speech-based controls ("turn back", "follow little/big"). These capabilities were demonstrated in moderately noisy environments where both the robot and operator were moving. Speech commands enabled control when the robot and operator were out of line of sight.

2.2.5.4 Adding a Handheld Pointing Device

Natural speech, used alone, can be ambiguous (e.g., "go over there") and needs some means of conveying details such as direction, either through further speech-based clarification, a PDA visual display-based map, a laser pointer, or a gesture, to address the ambiguity in deictic elements (e.g., "this", "that", "there", etc.) (Perzanowski et al. 1998, 2000a, 2000b).

3. Wearable Instrumented Systems

An alternative to camera-based systems for gesture-based controls is that of wearable instrumented systems. The most common technology used for gestures is instrumented gloves, but a few other approaches have been identified, as described in the following subsections. In this section, we describe the 2 main approaches to wearable gestural systems (e.g., instrumented gloves, electromyographic [EMG] sleeves) and approaches to gesture recognition using these systems. We then describe some military applications using these systems. While the focus is on robot control, we include some other applications, such as aircraft control and dismount communications. There are also applications developing within cockpit-type environments (Brown et al. 2011; DeVries et al. 2012; Slade and Bowman 2011); however, details are limited.

3.1 Wearable Instrumented Gloves

Instrumented gloves are the most common instantiation of wearable instrumented systems for robot control. The glove concept is congruent for many work situations where operators may already have to wear gloves. Early versions of these gloves were integrated for computer usage, in that the gloves could be used for computer interface actions such as menu selection. However, the reliance on a visual display was somewhat detrimental to performance (Kenn et al. 2007). For robot control, glove-based approaches are usually stand-alone, with the glove sending signals to robotic intelligence software for recognition, interpretation, and translation into computationally understandable and executable robotic behaviors.

Earlier instrumented gloves relied on sensors such as bend sensors, which react to changes of finger angles, and sensor electrodes (Karlsson et al. 1998). Others use touch sensors, magnetic trackers, embedded accelerometers, and electromagnetic position sensors with multiple degrees of freedom to convey information that is then mathematically interpreted. Optical fiber sensors have also been used to detect angular displacements of finger joints (Fujiwara et al. 2013). Iba et al. (1999) described the integration of a Cyberglove, a 6DOF position sensor to determine wrist position and orientation, a mobile robot, and a geolocation system that tracks

the robot location. The operator may give specific commands to the robot or wave in the direction the robot is to move (e.g., to the left or right). HMM algorithms are used for gesture recognition. Multiple users were used to train the 6 commands. Recognition under ideal conditions was very accurate (96%–100%), with fewer false positives associated when there was a “wait” state in the gesture, such that the gestures were more easily distinguished.

Barber et al. (2013) developed an instrumented glove combined with a 9-axis inertial measurement unit (IMU) for classification of 21 unique hand and arm gestures that were from the Army Field Manual (Headquarters, Department of the Army 1987) and modified Palm Graffiti. The wireless gesture-recognition glove incorporated bend sensors to distinguish pointing gestures and open/closed fist for determining the start/end of a gesture. They reported a 98% accuracy using a modified handwriting recognition statistical algorithm. The same algorithm was tested against a handheld device (Nintendo Wiimote), which also demonstrated an accuracy of 96% on the same set of gestures. Hill et al. (2015) demonstrated the use of this same system in an outdoor field assessment where arm and hand gestures were used to pause, resume, and direct control (e.g., move forward) a mobile robot.

While instrumented gloves in general have core traits in common, each approach to design has advantages and disadvantages specific to particular task demands. For example, bend sensors can be fragile in adverse environments or repeated use, and can only track static postures. Finger-touch sensors may move out of position over time. Accelerometers on fingertips and/or the back of the hand rely on precise finger movements and finger and hand position. Piezo sensors have also been used (Hu et al. 2009). While several types of gloves can accommodate static hand postures (positioning), dynamic hand and arm movements are more challenging or are not possible for effective gesture recognition by some of these instrumented systems. In short, the findings from one type of glove do not necessarily generalize to other types of gloves; they are not equal in capability or usability.

3.2 Other Wearable Sensors

While gloves are the more common approach to wearable sensors for gesture recognition, other types of wearable sensors have also been developed. Wu et al. (2010) used a 3-axis accelerometer mounted to the user’s wrist to record hand trajectories, which were then classified as 1 of 6 commands (“turn right/left”, “go straight”, “go back”, “rotate”, “stop”). They reported 92% accuracy using the DTW recognition algorithms. Lementec and Bajcsy (2004) used an array of wearable sensors to detect arm orientations. Yan et al. (2012) reported success with body-worn sensors (e.g., arm tape, wrist harness, thoracic and pelvic orientation sensors)

as well as with 3-D (i.e., Kinect) camera-based systems. While both the body-worn and camera-based systems were equally effective, there were additional constraints with the camera-based system beyond those of the body-worn sensors. For the camera-based systems, the controller must always face the camera and stay in a certain distance from the camera.

There is an emerging alternate approach for gesture control that is based on recognition of EMG signals, where EMG activity is recorded from forearm locations. In the field of prosthetics, EMG signals have been used for simple binary commands; however, advancements are enabling more complex coding based on postures. Crawford et al. (2005) described how they used this approach to control a robotic arm with gripper functions, from static hand postures. Electrodes were placed on 7 forearm locations and one on the upper arm. Features were extracted for each person (N = 3), for each of 8 hand postures, over 5 sessions. In each session, the subject maintained each of the 8 postures for 10 s. Robotic tasks ranged from simple (e.g., "move the arm to topple blocks") to more challenging (e.g., "pick up a designated object and drop in specified bin"). They reported classification accuracies over 90%. Progress has been reported with regard to adaptability to power source and reduction in muscle fatigue (Li et al. 2011) and more generalizable recognition of signals across multiple users (Matsubara and Morimoto 2013).

The JPL Biosleeve, developed at the Advanced Robotics Controls group at the Jet Propulsion Laboratory (California Institute of Technology) also uses this approach. Eight bipolar EMG sensors and an IMU) are mounted in a wearable sleeve that monitors forearm muscles. The Biosleeve development is focused toward National Aeronautics and Space Administration space missions to enhance telepresence control and for astronauts working side by side with robots (Assad et al. 2012; Wolf et al. 2013).

The concept is particularly apt for astronaut use outside the vehicle, as they are encumbered with heavy gloves, making traditional interfaces, such as joystick or touch interfaces, less effective. The sleeve is programmed to a particular user to develop recognition for static and dynamic gestures using algorithms to match temporal feature patterns to known templates. Recognition rates, based on 3 subjects, ranged from 93% to 99% over 17 static gestural commands. Recognition rates based on one subject resulted in 99% accuracy for 9 dynamic gestures.

3.3 Gesture Recognition by Instrumented Systems

The glove-based approaches have the same core challenge of gesture capture and coding as do the camera-based approaches. Gesture recognition analysis methods

for instrumented gloves overlap with approaches taken with camera-based systems: HMMs (Rabiner and Juang, 1986), FSM (Hong et al. 2000), DTW (Hu et al. 2009; Wu et al. 2010), and ANNs (Oz and Leu 2007). Hand and body gestures can be transmitted from a controller mechanism that contains IMU sensors to sense rotation and acceleration of movement. HMMs have the ability to model sequential information and have been used dominantly throughout the past decade (Ong and Ranganath 2005). In previous years, HMMs were used to recognize ASL with real-time color-based hand tracking (Starner and Pentland 1996). However, HMM has a relatively long training time with a large amount of training data. Others have found that the DTW approach is as effective as HMM, with a small set of gestures, while taking less time to prepare (Wu et al. 2010). Huang and associates (2011) describe a somewhat different approach to glove-based recognition, which abstracts a concept from raw data to recognize clusters of data. While the mathematic algorithms differ, there are still the same basic principles that underlie algorithmic recognition strategies to parse gestural movements into meaningful commands.

Earlier versions of wearable instrumented gloves relied on recognition of static hand gestures, as extrapolated by the glove sensors with regard to hand posture. Gesture recognition was based on specific features of a hand posture that may be somewhat artificial and difficult to replicate across different users. While intuitive in nature, and appreciated by users, the application of glove-based control can be challenging for robot control. Kenn et al. (2007) found that users would naturally look to the robot instead of their hands, thus allowing more attention to robot performance. However, the gestural commands were sometimes misinterpreted, and these misinterpretations had more negative consequence to robot control compared with other purposes, such as using gestures to control a presentation application. The negative consequences include robot movements that keep going if a stop command is not recognized, or robots that actuate in response to an unintended command. With human-robot situations, the tolerance for error may be very low.

Recent applications of the instrumented glove for robot control have been reported as successful. Boonpinon and Sudsang (2008) used a data glove to command multiple robots. More-recent technology adds the capability for more dynamic, movement-based gestures. IMU sensor technologies placed on the body provide an alternative approach to gesture recognition within uncontrolled environments. Iba et al.'s (1999) Cyberglove measures 18 joint angles of the fingers and wrist, with a 6DOF positioning system to determine wrist position and orientation. The system was used to recognize 6 gestures, based on a) hand opening, b) flat open hand, c) hand closing, d) index finger pointing, e) waving fingers to the left, f) waving fingers to the right, and g) none of the above. Kang et al. (2010) used a version of

the Cyberglove having a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis magnetometer. They developed a data fusion approach to integrate findings, consisting of 5 substeps. They reported an average of 94% recognition rate for 10 gestural commands. Rates were somewhat lower for 12 commands (91.7%) and 14 commands (86.5%). Jin and his associates (2011) used an instrumented glove orientation sensor to recognize static hand commands.

3.4 Military Applications: Instrumented Systems

3.4.1 Dismount Soldier Communications

Sullivan et al. (2010) described progress toward a DARPA-sponsored goal for a Soldier ensemble incorporating gesture recognition, head tracking, laser rangefinding, and augmented reality displays to support shared local situation awareness. Objects marked by one Soldier using gestures are portrayed to other Soldiers and shown as overlays in wearable visual displays. The gestural system included hand and upper arm gesture sensors that work with Soldiers wearing gloves.

Ellen et al. (2010) offer a concept of operations for use of wireless communication gloves to communicate, plan, and react while wearing Mission-oriented Protective Posture (MOPP) gear (i.e., protective ensemble for chemical and biologic threats) (Ceruti et al. 2009; Defense Tech Briefs 2010). While smartphone applications can be useful for Soldier use, touchscreen controls are not easily accomplished while wearing MOPP gear that includes heavy gloves. The gloves have magnetic and motion sensors for gesture recognition. Additional glove-based sensors would provide information regarding chemical and biological threats in the immediate area. Other sensors in this concept include GPS, physiological sensors, and video imagery. Researchers from this group also demonstrated the wireless glove for robot control (Tran et al. 2009). In this effort, glove-based sensors sent signals to a processing unit worn on the forearm, which then sent signals to a TALON robot. Commands were used to control robot position, camera, claw, and arm, such as “point camera” and “grab object”.

AnthroTronix has demonstrated (IMU-based hand and arm signal gesture recognition accuracy of 100% (Vice et al. 2001) via a custom instrumented glove interface. A more recent effort developed an instrumented glove for robot control and for hand and arm gestures for Soldier communications. Robot control was accomplished through navigation cues via static hand gestures (e.g., “move forward”, “move backward”, “turn left”, “turn right”, “stop”). Communication with other Soldiers was sent through hand and arm gestures (i.e., freeze, rally, danger, double time) that were received through signal patterns via a tactile vest. The tactile

vest allowed more immediate perception and understanding compared with standard visually communicated hand and arm signals (Elliott et al. 2014).

The US Army Research Laboratory and the University of Central Florida have demonstrated an IMU-based hand and arm signal gesture recognition system for interactions with an autonomous ground robot (Hill et al. 2015; Barber et al. 2015). The IMU-based instrumented glove recognized 8 dynamic hand and arm gestures to issue commands to the robot. Integrated within a multimodal interface that also used speech for verbal commands, the user could perform a gesture to pause or resume autonomous navigation of the robot in addition to simple drive commands (e.g., “move forward”, “move backward”, “turn left”).

Calvo et al. (2012) developed and evaluated a pointing device embedded within a tactical glove normally worn by US Air Force Battlefield Airmen during dismount military operations. Battlefield Airmen are the special operations force of the Air Force. The system detects the user’s wrist and finger movements through a 3-axis gyroscopic sensor. This capability was primarily explored for controlling cursor movements on a handheld device, as compared with using a touchpad or TrackPoint device. Moving the sensor with directional hand movements proportionally moves the cursor in the same direction. The user can also use their thumb to press buttons placed on the side of the index finger, to enable cursor movement, and perform cursor left clicks. While training time to reach asymptotic performance was longest with the glove than with the touchpad or TrackPoint device, throughput performance was significantly higher with the glove compared with the touchpad, which in turn, was significantly higher than the TrackPoint. Movement times were lowest with the glove; however, the glove was associated with higher error. This glove concept was compared to a handheld controller for micro-UAV control to navigate waypoints. Operators included 6 nonmilitary subjects performing in a simulation-based training environment. Waypoints were presented as red cubes in a 3-D virtual environment. Performance was not significantly different from the handheld controller. These results are quite promising and warrant further investigations using military subjects.

3.4.2 Robot Control

3.4.2.1 AnthroTronix

Figure 2 shows Soldiers using an instrumented glove for robot control and for hand-arm communications (Elliott et al. 2014).

Fig. 2 Soldier using instrumented glove for robot control (left) and communications (right)

3.4.2.2 SA Photonics

Researchers at the Space and Naval Warfare Systems Center (Ceruti et al. 2009; Tran et al. 2009) demonstrated application of a wireless communications glove for robot control and communications, along with other tasks, such as motion tracking, gesture recognition, data transmission, and reception, in normal and in extreme environments. They compared glove prototypes with and without a haptic (i.e., vibrotactile) capability for signal feedback and covert signal reception. Both types of gloves can be worn under a space suit or chemical protective gear and still operate effectively.

3.4.3 Aircraft Direction

Urban et al. (2004) used a motion tracker along with wearable sensors. Sensors were attached to armbands worn by the operator. Two sensors on each arm (i.e., upper and lower arm) were determined to be sufficient for most gestures in the Navy gesture lexicon for control of aircraft on the ground. While gestures were accurately recognized, there were problems associated with wired sensors (e.g., tangled wires, necessity of being near the base device) and fatigue associated with bulky sensors, indicating the need for smaller, wireless sensors.

Research and development of gestural communications for military use highlight positive potential for successful use and the limitations that must be addressed. The variation of gestural command approaches represents a corresponding variety of advantages and disadvantages. Identification of promising technology for a particular use should begin with consideration of purpose through analyses of operational requirements, task demands, and situational constraints.

3.4.4 Constraints in Military Operations

Some characteristics constrain the use of wearable sensors such as instrumented gloves. As identified in the following bullets, while some of the wearable instrumented systems can perform well in controlled environments, military operational environments may prove challenging for the current state-of-the-art technology.

- Operational range. The instrumented gloves/sensors must be used within system range. At this time, technology improvements to the electronic transmission capabilities (for sender and receiver) will be needed for longer distances.
- Latency. Latency of signals may be a problem when operational tempo is fast and movements quick.
- Multiple instrumented gloves/sensors. There may be networking limitations on the integration of multiple wearable controllers.
- Pointing gestures. Pointing has not yet been well-achieved while using wearable instrumented devices for gestures. Certainly part of the issue is that it is difficult to know what a pointing gesture is directed toward—what is being pointed at? Current technology may be able to be augmented with speech recognition.

4. Discussion

In our overview, we focused on hand and arm gestures, as they are most commonly used. For example, Mitra and Acharya (2007) identify 90% of gesture communication conveyed through hands and arms. Also, Soldiers already have familiarity of hand and arm signals for communication, some of which could also be considered as gestural commands to robots. The use of gestures is already common in military tactics and communication, as they have the advantage of being silent and easily used. Military hand and arm signals are documented in sources such as the US Army Field Manual No. 21-60 (Headquarters, Department of the Army 1987) and Marine Rifle Squad MWCP 3-11.2 (Headquarters, Department of the Navy 2002). In addition, certain mission scenarios may have further requirements, such as a need for covert operations (e.g., low noise and electronic transmissions) or combat operations that are characterized by high stress, high time pressure, high noise, low visibility, or night operations. Task demands of military ground robots will vary depending on mission requirements, such that the user may have to control robot movements, camera actions (e.g., pan, zoom, take pictures), or robot manipulations (e.g., grasping and manipulating objects).

The ease of use expected from gestures illustrates a core goal for effective use of gestural commands across many situations: that they be intuitive and easy to learn; ideally to the point of being discoverable (i.e., the user discovers the gestural command as a natural consequence of interaction). The best, most natural designs are those that are intuitive, that match the behavior of the system to the gestures humans would naturally use to accomplish the behavior, such as putting your hands under a faucet to turn it on, walking through a doorway, or pointing to direct attention to a location. Task context must also be a primary consideration in gesture evaluation: some gestures may be intuitive, easily learned, and performed for a single command but would be inappropriate for sustained repetitive use. Some situations will be more demanding on the user and the technology, such that each technology must be assessed in light of the intended use and users. However, general principles for gestural commands would include the following design goals to achieve high user acceptance, arising from perceived usefulness along with ease of use (Davis 1989):

- Easy to learn
- Easy to perform
- "Natural" movements
- Easy to remember
- Easily distinguishable gestures (from other gestures)
- Easily distinguished from normal work movements
- Easily distinguished from other people in surrounding area
- Easily distinguished from normal social gesticulations
- Easily recognized within a specified range (relevant to task)
- Socially acceptable
- On/off control
- Compatible with and supplemental to existing hand and arm gestures

While these characteristics seem self-evident, embedded within are issues with regard to measurement of these concepts. For example, concepts such as "intuitiveness" and "comfort" are subjective and perhaps somewhat ambiguous, although efforts have been made toward a more-quantitative approach to measurement (Stern et al. 2006). In addition, there are situational moderators, such that one solution does not fit all. For example, ease of learning and recognition can be significantly affected not only by the nature of the gesture(s) themselves, but

also by the number of gestures that are needed. Certainly, it is more difficult to learn, remember, and distinguish a large number of gestures. Comfort and ease of performance will also be greatly affected by gesture duration and use over time, and whether commands must be made while on the move. Social acceptance will also be affected by the environment; for example, speech-based controls would not be as appropriate for use in a quiet situation such as a library. Similarly, broad active gesticulations would not be appropriate for covert Soldier missions, such as an ambush. Designers must systematically analyze situation task demands to best tailor gestures (e.g., static hand postures, dynamic hand movements, etc.) with gestural recognition capabilities, environmental constraints, and need for precision (e.g., consequences of error or repetition). Given this, we turn attention to issues relevant to Soldier use and performance.

4.1 Issues Relevant to Soldier Use

It is evident that the field of gestural controls for robots has shown promise along a number of task demands, ranging from simple commands to more complex and autonomous tactical commands. Given that Soldier tasks can also vary, along with environmental demands and constraints, we identify the following mission and task-related factors for consideration when designing gestural controls for Soldier use.

4.1.1 Line of Sight/Distance between Robot and User

Many dismount Soldier missions are conducted with intact squads, where Soldiers are within line of sight. Some robot roles would normally take place in proximity to squad members, such as robotic mules to offload weight. While the constraints associated with the need for line of sight seems more applicable to camera-based systems, it can also be relevant to instrumented systems, because of the need for appropriate range for wireless signals. The need for line of sight or close range can affect the nature of gestures to be used. For example, larger, more whole-body dynamic movements may be more effective for camera-based systems. The need for direct line of sight with camera systems could affect military operations such as room clearing, reconnaissance, ordnance removal, and other situations where the robot may be distant and/or out of line of sight. Perhaps most important, the need for line-of-sight operations requires the Soldier-gesturer to be exposed, making the gesturer more vulnerable and potentially reducing survivability.

4.1.2 Visibility/Night Operations

For camera-based systems, the issues associated with the need for visibility is similar to that of line of sight and range. Camera-based systems that cannot be

effective in smoke, fog, mist, or night operations will be greatly limited in tactical applications. To be effective in these low-visibility conditions, camera-based systems will need to be augmented with specialized capabilities such as night- or thermal-vision capabilities. (Research and progress in this area warrant greater focus with regard to engineering and human factors). At this time, wearable instrumented systems in general have greater advantage for limited visibility operations.

4.1.3 Practicality of Using Special "Markers"

Some camera image recognition systems are augmented by the use of special clothing or wearable markers, which ease the processing complexity and load. Situational task demands will greatly affect whether this approach can be used. It should be noted that the special clothing or markers that make the user more visible and recognizable to the camera would not likely be suited for covert operations where Soldiers strive for discretion and secrecy. Special clothing/markers can make the Soldier a higher-risk target and more likely to be engaged by enemy forces, thus survivability may be compromised.

4.1.4 Practicality of Using Wireless Transmissions

Some operational missions may require silence, not only for audio, but also for electronic transmissions. Other missions may be more susceptible to jamming of such electronic transmissions. While camera systems are not vulnerable to jamming transmissions, the instrumented wearable systems would be rendered ineffective if transmissions were jammed.

4.1.5 Need for Deictic Information: Pointing Gestures

Deictic information depends on knowledge of situation context and personnel locations. The receiver robot must know where the operator is, as well as the direction of the pointing gesture. While pointing gestures are intuitive and perhaps the most useful of gestures, they are also quite difficult to instantiate successfully, for both camera-based and instrumented systems. If there is a need for deictic information, specifications should be generated with regard to the nature of the pointing gesture. For example, will the gesture indicate a general area? If so, how large is this area? Will the gesture be augmented by speech and object recognition (e.g., "bring me that ammo pack") or integrated with map-based information on a visual display? The multimodality of the entire system will greatly affect the utility of pointing gestures in general.

The effectiveness of any gesture set will be affected by the nature of its multimodal context. Gestures may be added to augment an existing system to increase

intelligibility in noisy environments. In the same way, speech can augment a vision-based gestural system when visibility is occluded. It is not likely at this time that a gestural command system would be completely stand-alone; instead, it will serve to augment and complement speech, keyboard, and or visual map displays. Scenario-based cognitive task analysis approaches including these issues should drive development of new Soldier concepts (Hoffman and Elliott 2010).

4.1.6 Ease of Producing and Sustaining Gestures

Situations should be analyzed with regard to the frequency and duration of gestural use. Gestures that may be easy to produce for a short period may become fatiguing over time. For example, robot controllers found that holding their hand steady and parallel to the ground was at first quite easy to accomplish but soon became fatigued after only 15 min. Thus, a relevant issue is whether the gestures have to be maintained continuously. This issue pertains to both camera-based and instrumented gestural systems.

4.1.7 Number of Gestures

The number of gestural commands that are needed for robot control will greatly affect the ease of use for the operator. Given a situation that requires more than 5–7 commands, particular attention must be given to gesture distinguishability, as well as ease of producing the gesture. As the number of gestural commands becomes higher, there should be greater consideration of gestures that use movement as well as static postures, including arm and/or whole body movements, to ease recognition. As the number of gestures increases, the training takes longer, the difference between gestures may be less than optimal, and the robot may not be able to determine differences at increased operating distances. This issue pertains to both camera-based and instrumented systems.

4.1.8 Number of Robots to Be Controlled by One Person

Gestures have been used to coordinate formation and movement of multiple robots. For example, an instrumented glove has been used to facilitate coordination of robot behaviors in multirobot systems having inherent decision rules to stay together, according to assigned distance parameters (Boonpinon and Sudsang 2008). In such a situation, the approach to gesture control is somewhat similar to the gestures and whistles used by herders to command sheep and/or herding dogs (e.g., the sheep have natural tendencies for formation and reaction to command stimuli [Phillips et al. 2012]). Multirobot situations must be carefully considered to match gestural commands to robot capabilities (e.g., level of autonomy and group formation

algorithms) and robot task demands, at both the individual and multiple-robot level of analysis. This issue pertains to both camera-based and instrumented systems.

4.1.9 Number of Coordinating Soldier Users

As robot-Soldier scenarios become more complex, there may be situations when the robot is considered a squad-level asset, much like any other squad member. In that case, careful attention must be given to provide not only the capabilities but the policies for squad-level use (i.e., the tactics, techniques, and procedures [TTPs] to be used by the Soldiers). When such TTPs are developed, the robot could, and should participate as an integral member of the squad, in training exercises (i.e., battle drills), to best ensure performance that is intuitive, efficient, and effective. In addition to consideration of TTPs, systems should be programmable for accepting unit standard operating procedures, which differ from unit to unit. This issue pertains to both camera-based and instrumented systems.

4.1.10 Combat Readiness

There are also the usual factors that must be considered for any Soldier system. These would include consideration of the type of power source, battery life, weight/bulk, sensitivity, accuracy, reliability, noise and light discipline, operational security, operational environment (sand/dust, rain, terrain, urban, etc.), durability, and maintainability.

In addition the system should be simple to operate, modular in design for specific application to various missions or tasks, designed to deny use by enemy against friendly, include design considerations for continuous operations missions lasting from 72–96 h, man-portable, air-droppable, and have decontamination procedures for nuclear, biological, and chemical environment (i.e., disposable vs. able to be decontaminated).

Table 2 lists primary considerations for gesture technology type that are relevant to use by Soldiers in operational contexts. The positive characteristics (Pros) and negative characteristics (Cons) of 2 types of gesture technology, camera-based systems and instrumented systems, are presented. Throughout the table, positive characteristics (Pros) for Soldier use are presented in plain font, while negative characteristics (Cons) for Soldier use are presented in italics.

Table 2 Considerations for gesture technology relevant to Soldier performance

Consideration	Camera-based system (Pros/*cons*)	Instrumented systems (Pros/*cons*)
Need for line of sight	• *Camera cannot capture and interpret command* • *If camera is too close, then it may miss part of the movement intended for command/control.* • *Trees, bushes, obstacles can severely restrict line of sight between operator and robot camera.* • *Operator must remain exposed in hostile environment to control robot*	• Gesture will be interpreted regardless of intermittent line of sight • Operator may be able to control robot from a covered or defilade position. • *Operator may be restricted based on dense foliage from executing proper gesture.*
Distance between robot and user	• *Camera may require close range for effective gesture recognition even when signaler is within line of sight.*	• *Dependent on network range capability (See also Wireless Transmissions below)*
Need for visibility during low-visibility operations	• May be addressed through specialized vision systems • *Daylight camera systems may not capture and interpret command in low visibility*	• Instrumented systems not affected by level of visibility
Need for wearable markers	• May be required for effective use in cluttered environments • *May enable targeting of operator, reducing safety and effectiveness*	• Not needed
Need for wireless transmissions	• Not needed	• *Depending on system, may be susceptible to interference/jamming*
Need for deictic (pointing) information	• May be augmented by speech, visual display, or pointing device (e.g., laser-pointing system) • *2-D camera systems have difficulty interpreting 3-D cue information*	• May be augmented by speech, visual display, or pointing device (e.g., laser-pointing system) • Pointing system may be incorporated in wearable or handheld systems

Table 2 Considerations for gesture technology relevant to Soldier performance (continued)

Consideration	Applies to BOTH camera and instrumented systems
Ease of producing and sustaining gestures	• Frequency and duration of gestures may impact fatigue and ability to hold gesture for some period of time
Number of gestures	• The number and distinguishability of gestures will greater affect the ease of use by Soldier
Number of robots to be controlled by one person	• Careful consideration needs to be made for multirobot systems.
Number of coordinating Soldier users	• For multirobot, multi-Soldier user scenarios, need careful attention to robot responses to gestures and policies for squad-level use.
Combat readiness	• Important factors for any Soldier system, including gesture technology, include power source, battery life, weight/bulk, operational security, operational environment, etc.

4.2 Future Directions

4.2.1 Integration with Speech

Given the relatively hands-free nature of gestural commands and constraints with regard to command vocabulary and syntax, it is likely that a combination with speech control would be beneficial. While control systems should be separable as an option (e.g., in circumstances where minimal sound is required, or one component is not working), it is clear that integration of multiple modalities will benefit from respective advantages. The combination of deictic gestures to support human-human interactions has been well established (Urban and Bajcsy 2005). Most efforts follow the structure of "put-that-there" system of Bolt (1980) that refers to object and locations by pointing and speaking. Research from neuroscience has argued that human gesture production is very much associated with language processing (Kelly et al. 2009; Mayberry and Jacques 2000) and are thus an intuitive pairing. Consistent with this view, Gullberg (1999) reported that over 50% of gestures used in spontaneous gesture production were to clarify speech utterances. Studies have shown demonstrable benefits from use of gestures to support deictic information, particular with regard to location and spatial relationships (St Clair et al. 2011b).

Speech-based controls have been developed with the goal of natural language interaction (Brooks et al. 2012). At this time, purely speech-based controls face a core challenge regarding communication of spatial relationships and explicit directions (e.g., "go to the east side of the third building behind the church"), and pointing gestures are expected to help clarify localization information. The challenge becomes even greater if one is commanding a robot with multiple degrees

of freedom in space; neither speech nor any programming language is well suited for these types of commands (Hirzinger 2001). Speech-based controls can also be problematic in noisy environments such as aircraft carriers. The addition of gestures to speech control systems should result in higher effectiveness based on complementary capabilities.

Marge et al. (2012) used a WoZ approach to investigate relative advantages of a traditional handheld robot controller with video feed, against alternatives that used speech or a combination of speech and gesture. The evaluation was based on a Soldier room-clearing mission, where the ground robot played the role of a fellow Soldier who guards the hallway and watches for enemy movement. Twenty-one of 30 participants were active duty in the military. In the traditional handheld display condition, the participant used a tablet computer with a virtual onscreen joystick, which also provided a video feed from the robot's camera. The participant was tasked to move and search rooms using a predefined route, look for and note certain objects in the environment, and monitor the video for passing people (indicated by a marker placed in front of the camera). In the speech condition, the video feed was replaced by a simulated capability where the robot alerts the participant through speech. In the combined speech and gesture condition, the speech replaced the video speech, and teleoperation of the robot was performed by a hand and arm gesture to stop and start robot "follow" movements. Actual commands were accomplished through experimenter control of robot movements, allowing a more controlled investigation of options independent of other performance factors (e.g., capability and reliability of speech and gesture control). Results indicated that significantly faster speed and lower workload associated with the speech-gesture combined display, particularly for the military participants, who had less experience with robot controllers. Other participants were the robot developers and technicians who had more experience with the handheld controller. Participants expressed higher preference for the speech-gesture combination, as they allow heads-up hands-free control, allowing more attentional resources to their surroundings. While not conclusive, given the WoZ approach, results are promising and support further development of actual capability.

Sofge and colleagues discuss their application of 2 cognitive architectures (i.e., ACT-R and Polyscheme) to achieve understanding of natural language commands that involve spatial reasoning and perspective-taking (Sofge et al. 2003b). Related efforts are also focused on the modeling of spatial reference in natural language (Adams and Skubic 2005; Blisard and Skubic 2005; Luke et al. 2005), and the integration of speech with pointing gestures (St Clair et al. 2011b). Typically, the pointing gesture is used in conjunction with speech to accomplish labeling of objects (e.g., "this is saucer1"), where the pointing gesture assists in clarifying

“which” object. Pointing gestures are also used to indicate a location (e.g., “move 50 m over there” [Brooks and Breazeal 2006]). Kennedy and Rybski (2007) reported the integration of vision-based human/limb detection and tracking with the capability for a human to teach a robot how to do tasks through speech and physical demonstration. More basic research investigates the natural use of speech and gestures for human-to-human dialogs pertaining to spatial relationships (Lucking et al. 2012). As robots become more capable (e.g., intelligent, autonomous), the guidelines pertaining to human-human communications will become more relevant. It is sensible to conclude that systems using both speech and gestures will naturally evolve towards more effective and intuitive robot control systems.

Stiefelhagen et al. (2004) developed a speech-gesture control system as a natural interaction with an assistive robot, in a kitchen scenario. Speech, head pose, and gestures were integrated with visual recognition technology to communicate commands such as “what is in the refrigerator”, “please set the table”, “please turn off/on the light”, and “please bring me a certain object”. The system included a JANUS speech recognition system developed at the University of Karlsruhe, 3-D face and hand tracking, speech synthesis, and a stereo camera system with pan/tilt. Remote microphones were used in lieu of head-mounted alternatives, to minimize user discomfort. Depth information allowed gesture recognition that was more robust to lighting changes. Head pose information was used to signify direction and to determine whether the user was directing speech to the robot as a command. When combined with pointing gestures, the head pose information increased accuracy of interpretation, by reducing false positive error rate from 26% to 13%. The robot visual recognition system was programmed to recognize objects such as cups, dishes, forks, knifes, spoons, and lamps. Thus, combination of speech, gesture, and head pose information allowed interpretation of ambiguous phrases such as “get me that fork” or “switch on that lamp”. However, performance data for the system were sparse and restricted to performance of each component rather than the system as a whole. The system was described as a work in progress, with current goals toward a more humanoid robot with 2 arms. Similarly, Rogalla and associates (Rogalla et al. 2002) also reported progress on integrated vision-based recognition of gestures and objects with speech-based control, for assistive tasks. In their approach, emphasis was on hand silhouette recognition of hands with color segmentation of objects, combined with “ViaVoice” speech recognition capability, to accomplish tasks such as “take the cup from the table in front of you”.

While benefits have been demonstrated, challenges remain with regard to effective integration of speech and gesture, particularly in multi-object environments in a 3-D world. These more complex scenarios represent a complex and unstructured problem (Brooks and Breazeal 2006). Similar issues are faced as researchers strive

to develop an interface that integrates head-mounted visual display, speech, and gesture controls for the commander as he or she is seated within a moving command vehicle (Neely et al. 2004). Gestures are most naturally effective in situations of physical co-presence, where the robot and the operator can establish a joint visual understanding of the environment, with physical and directional referents. This is particularly true if robot recognition of gestures is dependent on a camera-based system. In a somewhat different approach, (Taylor et al. 2014) used smartphone technology attached to a user's wrist to capture both speech and gestural movements, which was sent to a remote laptop for processing and command of a robotic mule. It is clear that many approaches are being explored from different perspectives to more fully achieve effective and naturalistic integration of speech and gesture.

4.2.2 Integration with Handheld Devices

The smartphone is a core element of the Army concept of operations for the ground Soldier (Barker 2013), providing the opportunity to use numerous apps to support mission tasks. While the concept for Soldier use is a popular one, with a dedicated program of effort, many challenges remain to be addressed to achieve capabilities typical of civilian use (Erwin 2011) due to limitations associated with secure use. Smartphone applications have incorporated 3-D audio and tactile feedback for waypoint navigation for US Air Force Battlefield Airmen (Calvo et al. 2013) and for robot control, incorporating features such as vision-based recognition and speech control (Checka 2011). The smartphone platform can also be used to support gestural commands. Two types of handheld devices are prevalent with regard to gesture control. One would be the use of the device to assist in gesture recognition, by being the object that is recognized. The Kinect gaming device is such an example. The user holds the device, which is tracked by a camera system. This approach was shown to be easy to learn for simple navigation commands to a robot, so that the device is used like a virtual leash (Olufs and Vincze 2009). The XWand, developed by Microsoft Research, is a wand-like device that enables the user to point at an object and control it using gestures and/or voice. For example, lights and music can be turned on and off by pointing to the device (light switch, music player) and saying "lights on" or "volume up" (Wilson and Shafer 2003). The Nintendo Wii remote controller has also been used as a gestural device to send 7 different communications, representative of Army hand and arm signals to a tactile belt (Varcholik and Merlo 2008).

Smartphone visual displays can also be used to augment speech or gesture. A map-based display addresses the challenge of directing a robot to a particular location or object. The visual display may be used with touch gestures, or integrated with

speech-based specifications. A grid-based map display allows the operator to refer to a grid-based location when directing the robot. Alternatively, an object-based approach is based on the labeling of referent objects, which can be referred to from different vantage points over time (Walter et al. 2010; Pettitt et al. 2014). Gopalakrishnan et al. (2005) demonstrated gains achieved through integrating camera-based gestures with visual displays, laser-based localization, and speech recognition capabilities.

Pointing devices have also been used to enhance the precision of the gesture to point to areas or objects. Kemp and his associates (Kemp et al. 2008) utilized an off-the-shelf green laser pointer, integrated with an omnidirectional catadioptric system (i.e., an optic combining reflection and refraction, such as lenses and curved mirrors) with a narrowband green filter. The user points at the object of interest, which is located by the robot. They reported 99.4% accuracy with regard to the robot looking at the correct object and estimating its 3-D location, and in 90% of the trials the robot successfully moved to the object and picked it up. Objects were within 3 m. Patel and Abowd (2003) developed a 2-way laser-assisted capability on a cell phone, which can select and communicate with photosensitive tags placed in the environment. This general approach is promising in that it avoids the complexity with regard to intelligent understanding of spatial relationships and recognition of objects and/or pointing gestures.

One recommendation with regard to handheld objects in general is to make the device easy to find if misplaced. In particular, this can be very important for Soldiers in military operations, given combat missions that are often executed at night, by users under high stress. A simple GPS chip within the device can indicate its position on a map. In addition, the device might emit audio cues when requested. In the same way, the robot itself may be beyond line of sight and in need of retrieval. In that case, a tactile belt display can respond to GPS signals and guide the wearer to the robot location, while leaving the hands and eyes free, to attend to weapons and surroundings (Pomranky-Hartnett et al. 2015).

Some characteristics that constrain the use of handheld devices and should be considered prior to development or selection for use are as follows:

- Operational range. The handheld devices must be used within system range. At this time, technology improvements to the electronic transmission capabilities (for sender and receiver) will be needed for longer distances.
- Latency. Latency of signals may be a problem when operational tempo is fast and movements quick.

- Multiple handhelds. There may be networking limitations on the integration of multiple handheld controllers.
- Ability to be hands free. By their very nature, handheld devices are not hands free. Even if stored in a pocket, a hand must be used to hold the device for gesture control.

5. Conclusions

It is clear that much progress has occurred with regard to development of gestural command systems, and that progress is ongoing. In addition, there is great variety of technology approaches. In this report, we describe many of these approaches, and offer an organizing framework that can allow the developer to more closely consider situational context and task demands to better identify the technology most suited to the purpose at hand.

It is also clear that integration with other modalities (e.g., visual map displays, speech control, and bidirectional communication) will offer a wider range of applications and greater effectiveness with regard to speed, accuracy, and ease of use. While camera-based systems are currently limited, research is ongoing to enable the camera system to more closely approximate the capability to not only "see", but to understand and interpret, not only gestures but situational context. At this time, however, both camera-based and instrumented systems have limitations that must be considered when choosing the most appropriate system for the task and situational demands at hand. For military use, generalizable recommendations would include weight, bulk, maintainability, power consumption, and ease of use. In addition, the use of wearable networked systems will always present security issues (Hudgens 2013). Proof-of-concept technology must address these issues before transition to combat situations.

6. References

Abidi S, Williams M, Johnston B. Human pointing as a robot directive. Proceedings of the 8th ACM/IEEE International Conference on Human Robot Interaction; 2013 Mar 4–6; Tokyo, Japan. Piscataway (NJ): IEEE Press; c2013. p. 67–68.

Ablavsky V. Taxiing via intelligent gesture recognition. Lakehurst (NJ): Naval Air Systems Command; 2004 Feb 4. Contract N00014-03-M-0327.

Adams J, Skubic, M. Introduction to the special issue on human-robot interaction. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans. 2005;35(4):433–437.

Aghajan H, Wu C. Layered and collaborative gesture analysis in multicamera networks. Proceedings of the 2007 IEEE International Conference on Acoustics. Speech, and Signal Processing (ICASSP); 2007 Apr 15–20; Honolulu, HI. ADM002013.

Alon J, Athitsos V, Sclaroff S. Simultaneous localization and recognition of dynamic hand gestures. Proceedings of the Seventh IEEE Workshops on Application of Computer Vision, WACV/MOTIONS ′05; 2005 Jan 5–7, Breckenridge, CO. Piscataway (NJ): IEEE Press; c2005. Vol 1 p. 254–260.

Assad C, Wolf M, Stoica A, Theodoridis T, Glette K. BioSleeve: A natural EMG-based interface for HRI. Proceedings of the 8th ACM/IEEE International Conference on Human Robot Interaction; 2012 Mar 3–6, Tokyo, Japan. Piscataway (NJ): IEEE Press; c2012. p. 69–70.

Axe D. War bots: US military robots are transforming war in Iraq, Afghanistan, and the future. Ann Arbor (MI): Nimble Books, LLC; 2008.

Baker J. The DRAGON system: An overview. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1975;23:24–29.

Barattini P, Morand C, Robertson N. A proposed gesture set for the control of industrial collaborative robots. Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication; 2012 Sep 9–13; Paris, France.

Barber DJ, Abich J, Phillips E, Talone AB, Jentsch F, Hill SG. Field assessment of multimodal communication for dismounted human-robot teams. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2015 Oct 26–30; Los Angeles, CA. Thousand Oaks (CA): SAGE Publications. 2015;59(1):921-925.

Barber D, Lackey S, Reinerman-Jones L, Hudson I. Visual and tactile interfaces for bi-directional human robot communication. In: Karlsen RE, Gage DW, Shoemaker CM, Gerhart GR, editors. Proc. SPIE 8741, Unmanned Systems Technology XV; 2013 Apr 29–May 3; Baltimore, MD. Bellingham (WA): International Society for Optics and Phonics; c2013. doi:10.1117/12.2015956.

Barker J. Operational test agency assessment report for the Nett warrior system. Aberdeen Proving Ground (MD): Army Evaluation Center (TEAE-MG) (US); 2013.

Basanez L, Suarez R. Teleoperation. In: Nof S, editor. Springer Handbook of Automation Part C. Heidelberg (Germany): Springer-Verlag Berlin Heidelberg; 2009. p. 449–468.

Beach G, Cohen C. Recognition of a computer based human gestures for device control and interacting with virtual worlds. Ann Arbor (MI): Cybernet Systems Corp.; 2001 Apr 23. Report No.: CSC-01-316-D. ADB265922.

Becker M, Kefalia E, Mael E, von der Malsburg C, Pagel M, Triesch J, Vorbruggen J, Wurtz R, Zadel S. Gripsee: a gesture controlled robot for object perception and manipulation. Autonomous Robots. 1999;6:203–221.

Beer J, Prakash A, Smarr C, Mitzner T, Kemp C, Rogers W. Commanding your robot: older adults' preferences for methods of robot control. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2012 Oct 22–26, Boston, MA; 2012;56(1):1263–1267.

Blisard N, Skubic M. Modeling spatial referencing language for human robot interaction. Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication; 2005 Aug 13–15; Nashville, TN. Piscataway (NJ): IEEE Press; c2005. p. 698–703.

Bolt R. Put-that-there: voice and gesture at the graphics interface. ACM Computer Graphics. 1980;14(3):262–270.

Boonpinon N, Sudsang A. Formation control for multi-robot teams using a data glove. 2008 IEEE International Conference on Robotics, Automation and Mechatronics; 2008 Sep 21–24; Chengdu, China. Piscataway (NJ): IEEE Press; c2008. p. 525–531.

Bremner P, Pipe A, Melhuish C, Fraser M, Subramanian S. Conversational gestures in human-robot interaction. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics; 2009 Oct 11–14; San Antonio, TX. Piscataway (NJ): IEEE Press; c2009. p. 1645–1649.

Brooks A, Breazeal C. Working with robots and objects: revisiting deictic reference for achieving spatial common ground. Presented at the 2006 Conference on Human Robot Interaction; 2006 Mar 2–3; Salt Lake City, UT.

Brooks D, Lignos C, Finucane C, Medvedev M, Perera I, Raman V, Kress-Gazit H, Marcus M, Yanco H. Make it so: continuous, flexible natural language interaction with an autonomous robot. Proceedings of the AAAI 2012 Workshop on Grounding Language for Physical Systems; 2012 Jul 22–23; Toronto, Ontario, Canada.

Brooks R. Natural tasking of robots based on human interaction cues. Cambridge (MA): Massachusetts Institute of Technology Cambridge Lab for Computer Science; 2005. Technical Report ADM001779.

Brown M, Mellott K, Grant N. Digital flight gloves. Wright-Patterson Air Force Base (OH): Air Force Research Laboratory (US); 2011. DTIC report No.: ADB380437.

Brown R. The infantry squad: decisive force now and in the future. Military Review. 2011 Nov–Dec 2–9. [accessed 2016 June 21]. http://usacac.army.mil/CAC2/MilitaryReview/Archives/English /MilitaryReview_20111231_art004.pdf.

Calvo A, Burnett G, Finomore V, Perugini S. The design, implementation, and evaluation of a pointing device for a wearable computer. Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting; 2012 Oct 22–26; Boston, MA. Thousand Oaks (CA): SAGE Publications. 2012;56(1):521–525.

Calvo A, Finomore V, Burnett G, McNitt T. Evaluation of a mobile application for multimodal land navigation. Proceedings of the 57th Annual Meeting of the Human Factors and Ergonomics Society; 2013 30 Sep–4 Oct; San Diego, CA. Thousand Oaks (CA): SAGE Publications. 2013;57(1):1997–2001.

Celik I, Kuntalp M. Development of a robotic arm controller by using hand gesture recognition. 2012 International Symposium on Innovations in Intelligent Systems and Applications; 2012 July 2–4; Trabzon, Turkey. Piscataway (NJ): IEEE Press; c2012. p. 1–5.

Ceruti M, Dinh VV, Tran NX, Phan HV, Duffy LT, Ton T-A, Leonard G, Medina E, Amezcua O, Fugate S et al. Wireless communication glove apparatus for motion tracking, gesture recognition, data transmission, and reception in extreme environments. Proceedings of the 2009 ACM International

Symposium on Applied Computing; 2009 Mar 9–12; Honolulu, HI. New York (NY): ACM; c2009. p. 172–176.

Checka N. Handheld apps for warfighters. Arlington (VA): Defense Advanced Research Projects Agency; 2011 Oct 24. Contract No.: W31P4Q-11-C-003, Doc ID 1281.

Checka N, Demirdjian D. Automated analysis and classification of anomalous 3D human shapes and hostile actions. Wright-Patterson AFB (OH): Air Force Research Laboratory (US); 2010. Technical Report No.: AFRL-RH-WP-TR-2010-0151.

Cheng L, Sun Q, Su H, Cong Y, Zhao S. Design and implementation of human robot interactive demonstration system based on Kinect. Proceedings of the 2012 24th Chinese Control and Decision Conference (CCDC); 2012 May 23–25, Taiyuan, China. Piscataway (NJ): IEEE Press; c2012. p. 971–975.

Chian T, Kok S, Lee G, Wei O. Gesture-based control of mobile robots. Proceedings of the 2008 IEEE Conference on Innovative Technologies in Intelligent Systems and Industrial Applications; 2008 Jul 12–13; Cyberjaya, Malaysia.

Choi C, Ahn J, Byun H. Visual recognition of aircraft marshalling signals using gesture phase analysis. IEEE Intelligent Vehicles Symposium; 2008 Jun 4-6; Eindhoven, The Netherlands.

Cohen C. An automatic learning gesture recognition interface for dismounted soldier training systems. Orlando (FL): Naval Air Warfare Center Training Systems Division; 2000. Final report.

Cohen C. A control theoretic method for categorizing visual imagery as human motion behaviors. 34th Applied Imagery Pattern Recognition Workshop; 2005 Oct; Washington, DC.

Cohen C. Motion tracking, gesture recognition, and surveillance: capability summary. Ann Arbor (MI): Cybernet Systems Corporation; 2013 Mar 13.

Colasanto L, Suarez R, Rosell J. Hybrid mappinVog for the assistance of teleoperated grasping tasks. IEEE Transactions on Systems, Man, and Cybernetic Systems. 2013;43(2):2168–2216.

Corradini A, Gross H. Camera-based gesture recognition for robot control. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks 2000; (IJCNN 2000), Vol 6; 2000 July 24–27; Como, Italy. Los Alamitos (CA): IEEE Computer Society Press; c2000. p. 133–138.

Crawford B, Miller K, Shenoy P, Rao R. Real-time classification of electromyographic signals for robotic control. Proceedings of AAAI-05, Association for the Advancement of Artificial Intelligence; 2005 July 9–13; Pittsburgh, PA. Palo Alto (CA): AAAI Press; 2005. p. 523–528. http://www.aaai.org/Papers/AAAI/2005/AAAI05-082.pdf.

Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly. 1989;13(3):319–314.

Defense Tech Briefs. Gesture-directed sensor information fusion gloves: These wireless gloves provide communications capabilities for warfighters in hazardous environments. San Diego (CA): Space and Naval Warfare Systems Center Pacific; 2010 Apr.

DeVries S, Nightenshelser S, Thogerson P. Wolf digital smart glove system. Wright Patterson AFB (OH): Air Force Research Laboratory (US); 2012. Technical Report No.: RH-WP-TR-2012-0017.

Ellen J, Ceruti M, Medina E, Duffy L. Gesture-directed sensor-information fusion for communication in hazardous environments. Proceedings of the 14th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI); 2010 29 June–2 July; Orlando, FL. Vol. II p. 116–119.

Elliott LR, Skinner A, Pettitt R, Vice J. Utilizing glove-based gestures and a tactile display for soldier covert communications and robot control. Aberdeen Proving Ground (MD): Army Research Laboratory Human Research and Engineering Directorate (US); 2014 Jun. Technical Report No.: ARL-TR-6971.

Elmezain M, Al-Hamadi A, Appendrodt J, Michaelis B. A hidden Markov model-based continuous gesture recognition system for hand motion trajectory. Proceedings of the 19th International Conference on Pattern Recognition (ICPR); 2008 Dec 8–11; Tampa, FL. Piscataway (NJ): IEEE Press; c2008. p. 1–4.

Erwin S. Smartphones for soldiers campaing hits wall as Army experiences growing pains. National Defence. 2011;95:691.

Fujiwara E, Miyatake D, Ferreira M, Suzuki C. Development of a glove-based optical fiber sensor for applications in human robot interaction. Proceedings of the 2013 8th ACM/IEEE International Conference on Human Robot Interaction; 2013 Mar 3–6; Tokyo, Japan. Piscataway (NJ): IEEE Press; c2013. p. 123–124.

Garg P, Aggarwal N, Sofat S. Vision based hand gesture recognition. World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering. 2009;3(1):186–191.

Gavrilla D. The visual analysis of human movement: a survey. Proceedings of Computer Vision and Image Understanding. 1999;73(1):82–98.

Goodrich M, Schultz A. Human-robot interaction: a review. Foundations and Trends in Human-Computer Interaction. 2007;1(3):203–275.

Gopalakrishnan A, Greene S, Sekmen A. Vision-based mobile robot learning and navigation. ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication; 2005 Aug 13–15; Nashville, TN. Piscataway (NJ): IEEE; c2005. p. 48–53.

Gordon G, Chen X, Buck R. Person and gesture tracking with smart stereo cameras. In: Corner BD, Mochimaru M, Sitnik R. Proc. SPIE 6805, Three-Dimensional Image Capture and Applications; 2008 Jan 27; San Jose, CA. Bellingham (WA): International Society for Optics and Phonics; c2008. doi:10.1117/12.759206.

Gullberg M. Communication strategies, gestures, and grammar. Acquisition et Interaction en Langue Etrangere. 1999;2:61–71.

Hashiyama T, Sada K, Iwata M, Tano S. Controlling an entertainment robot through intuitive gestures. Proceedings of the 2006 IEEE International Conference on Systems, Man, and Cybernetics; 2006 Oct 8–11; Taipei, Taiwan. Piscataway (NJ): IEEE; c2006. p. 1909–1914.

Hassanpour R, Wong S, Shahbahrami A. Vision-based hand gesture recognition for human computer interaction: a review. Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction; 2008 Jul 25–27; Amsterdam, The Netherlands.

Headquarters, Department of the Army. Visual signals. Chapter 2: arm and hand signals for ground forces. Washington (DC): Headquarters, Department of the Army; 1987. Field Manual No. FM 21-60.

Headquarters, Department of the Navy. Marine rifle squad. Quantico (VA): Marine Corps Combat Development Command (US); 2002. Fleet Marine Force Manual MCWP 3-11.2.

Hill SG, Barber D, Evans AW. Achieving the vision of effective Soldier-robot teaming: recent work in multimodal communication. HRI'15. Proceedings of

the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction – Extended Abstracts; 2015 Mar 2–5, Portland, OR. New York (NY): ACM; 2015. ACM 978-1-4503-3318-4/15/03. http://dx.doi.org/10.1145/2701973.2702026.

Hirzinger G. On the interaction between human hand and robot – from space to surgery. Proceedings of the 2001 IEEE International Workshop on Robot and Human Interactive Communication; 2001 Sep 18–21; Paris, France. Piscataway (NJ): IEEE Press; c2001. p. 1.

Hoffman R, Elliott L. Using scenario-based envisioning to explore wearable flexible displays. Ergonomics in Design. 2010;18(4):4–8.

Hong P, Turk M, Huang, T. Gesture modeling and recognition using finite state machines. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition; 2000 Mar 28–30; Grenoble, France. Piscataway (NJ): IEEE Press; c2000. p. 410–415.

Hu J, Wang J, Sun G. Self-balancing control and manipulation of a glove puppet robot on a two-wheel mobile platform. Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2009 Oct 11–15; St. Louis, MO.

Huang Y, Monekosso D, Wang H, Augusto J. A concept grounding approach for glove-based gesture recognition. Proceedings of the 2011 7th International Conference on Intelligent Environments; 2011 July 25–28; Nottingham, UK. Piscataway (NJ): IEEE Press; c2011. p. 358–361.

Hudgens J. Wearable devices and security implications. Fort Belvoir (VA): Army G2X (US); 2013. Technical Report No.: SURVIAC-TR-14-1311.

Iba S, Paredis C, Khosla P. Intention aware interactive multi-modal robot programming. Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2003 Oct 27–31; Las Vegas, NV.

Iba S, Weghe M, Paredis C, Khosla P. An architecture for gesture-based control of mobile robots. Proceedings of the 1999 IEEE/RSI International Conference on Intelligent Robots and Systems; 1999 Oct 17–21; Kyongjui, South Korea. Piscataway (NJ): IEEE Press; c1999. p. 851–857.

Jin S, Li Y, Lu G, Luo J, Chen W, Zheng X. Som-based hand gesture recognition for virtual interactions. Proceedings of the 2011 IEEE International Symposium on VR Innovation (ISVRI); 2011 Mar 19–23; Singapore. Piscataway (NJ): IEEE Press; c2011. p. 317–322.

Jones C. Tactical teams: cooperative robot/human teams for tactical maneuvers. Redstone Arsenal (AL): Army Aviation and Missile Command (US)/Defense Advanced Research Projects Agency (US); 2007. ARPA order U051-48.

Jung H, Ou S, Grupen R. Learning to perceive human intention and assist in a situated context. Workshop on learning for human-robot interaction modeling, robotics: science and systems (RSS); 2010 Jun; Zaragoza, Spain.

Kang S, Kim J, Lee J. Intuitive control using a mediated interface module. International Conference on Advances in Computer Entertainment Technology; 2009 Oct; Athens, Greece. ADM002330.

Kang S, Kim J, Sohn J. A robot control approach based on multi-level data fusion. Proceedings of the 2010 International Conference on Control, Automation, and Systems; 2010 Oct 27–30; Gyeonggi-do, Korea. Piscataway (NJ): IEEE Press; c2010. p. 1777–1780.

Kania R, Del Rose M. Autonomous robotic following using vision based techniques. Warren (MI): Army Research Development and Engineering Command, Visual Perception Laboratory (US); 2005. Technical Report No.: ADA439387.

Karlsson N, Karlsson B, Wide P. A glove equipped with finger flexion sensors as a command generator used in a fuzzy control system. Proceedings of the 1998 IEEE Instrumentation and Measurement Technology Conference; 1998 May 18–21; St Paul, MN. Piscataway (NJ): IEEE Press; c1998. p. 441–445.

Kelly S, Ozurek A, Maris E. Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. Psychological Science. 2009;21(2):26–267.

Kemp C, Anderson C, Nguyen H, Trevor A, Xu Z. A point and click interface for the real world: laser designation of objects for mobile manipulation. Proceedings of the Human Robot Interaction HRI International Conference; 2008 Mar 12–15; Amsterdam, Netherlands. Piscataway (NJ): IEEE; c2008. p. 241–248.

Kendon A. Gesture: Visible action as utterance. Cambridge (UK): Cambridge University Press; 2004.

Kenn, H, van Megen F, Sugar R. A glove-based gesture interface for wearable computing applications. Proceedings of the 4th International Forum on Applied Wearable Computing; 2007 Mar 12–13; Tel Aviv, Israel. Piscataway (NJ): IEEE Press; c2007. p. 1–10.

Kennedy W, Bugajska M, Marge M, Adams W, Fransen B, Perzanowski D, Schultz A, Trafton G. Spatial representation and reasoning for human-robot collaboration. Proceedings of the 22nd Conference on Artificial Intelligence; 2007 Jul 22–26; Vancouver, Canada. Palo Alto (CA): AAAI Press; c2007. p. 1554–1559.

Kennedy W, Rybski P. Peer to peer embedded human robot interaction. Amherst (NH): MobileRobots Inc.; 2007. ADB331450.

Khan R, Ibraheem N. Survey on gesture recognition for hand image postures. International Journal of Artificial Intelligence and Applications. 2012;3(4):161–174.

Kobayashi T, Haruyama S. Partly-hidden Markov model and its application to gesture recognition. Acoustics, Speech, and Signal Processing; Proceedings of 1997 IEEE International Conference; 1997 Apr 21–24; Munich, Germany. Piscataway (NJ): IEEE; c1997. p. 3081–3084 vol 4.

Konda K, Konigs A, Schulz H, Schulz D. Real time interaction with mobile robots using hand gestures. Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction; 2012 Mar 5–8; Boston, MA. Piscataway (NJ): IEEE; c2012. p. 177–178.

Kortenkamp D, Huber E, Bonasso P. Recognizing and interpreting gestures on a mobile robot. Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96); 1996 Aug 4–8; Portland, OR. Palo Alta (CA): AAAI; c1996.

Lai K, Konrad J, Ishwar P. A gesture-driven computer interface using Kinect. Proceedings of 2012 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI); 2012 Apr 22–24; Santa Fe, NM. Piscataway (NJ): IEEE; c2012. p. 185–188.

Lambrecht J, Kleinsorge M, Kruger J. Markerless gesture-based motion control and programming of industrial robots. Proceedings of the IEEE 16th Conference on Emerging Technologies and Factory Automation; 2011 Sep 5–9; Toulouse, France. Piscataway (NJ): IEEE; c2011. p. 1–4.

Lathuiliere F, Herve J. Visual hand posture tracking in a gripper guiding application. Proceedings of the 2000 IEEE International Conference on Robotics and Automation; 2000 Apr 24–28; San Francisco, CA. Piscataway (NJ): IEEE; c2000. p. 1688–1694.

Lementec J, Bajcsy P. Recognition of arm gestures using multiple orientation sensors: gesture classification. In: Proceedings of the 7th IEEE International

Conference on Intelligent Transportation Systems; 2004 Oct 3–6; Washington, WA. Piscataway (NJ): IEEE; c2004. p. 965–970.

Li L, Looney D, Park C, Rehman N, Mandic D. Power independent EMG-based gesture recognition for robotics. Proceedings of the 23rd Annual International Conference of the IEEE EMBS; 2011 Aug 30–Sep 3; Boston, MA.

Li M, Pan W. Gesture-based robot's long-range navigation. Proceedings of the 7th International Conference on Computer Science and Education; 2012 Jul 14–17; Melbourne, Australia. Piscataway (NJ): IEEE Press; c2012. p. 996–1001.

Lii N, Chen Z, Roa M, Maier A, Pleintinger B, Borst C. Toward a task space framework for gesture commanded telemanipulation. Proceedings of the 2012 IEEE RO-MAN: 21st IEEE International Symposium on Robot and Human Interactive Communication; 2012 Sep 9–13; Paris, France. Piscataway (NJ): IEEE; c2012. p. 925–932.

Lilley A. Army expeditionary warrior experiment (AEWE) spiral H final report. Aberdeen Proving Ground (MD): Army Test and Evaluation Command (CSTE-AEC-FFE) (US); 2013.

Littman E, Drees A, Rittern H. Robot guidance by human pointing gestures. Proceedings of the International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing; 1996 Aug 21–23; Venice, Italy. Piscataway (NJ): IEEE; c1996. p. 449-457.

Lucking A, Bergman K, Hahn F, Kopp S, Rieser H. Data-based analysis of speech and gesture: The Bielefeld speech and gesture alignment corpus (SaGA) and its applications. Journal of Multimodal Interfaces. 2012 Aug 12. [accessed 2016 May 18] DOI 10.1007/s12193-012-0106-8.

Luke R, Keller J, Skubic M, Senger S. Acquiring and maintaining abstract landmark chunks for cognitive robot navigation. Proceedings of the IEEE International Conference on Intelligent Robots and Systems; 2005 Aug 2–6; Alberta, Canada. Piscataway (NJ): IEEE; c2005. p. 2566–2571.

Malima A, Ozgur E, Cetin M. A fast algorithm for vision-based hand gesture recognition for robot control. Proceedings of the 2006 IEEE 14th Conference on Signal Processing and Communications Applications; 2006 Apr 17–19; Antalya, Turkey. Piscataway (NJ): IEEE; c2006. p. 1–4.

Manigandan M, Jackin M. Wireless vision-based mobile robot control using hand gesture recognition through perceptual color space. Proceedings of the 2010 International Conference on Advances in Computer Engineering; 2010 Jun 20–21; Bangalore, India. Piscataway (NJ): IEEE; c2010. p. 95–99.

Marge M, Powers A, Brookshire J, Jay T, Jenkins O, Geyer C. Comparing heads-up, handsfree operation of ground robots to teleoperation. In: Durrant-White H, Roy N, Abbeel P, editors. Robotics: science and systems. Cambridge (MA): MIT Press; 2012.

Masum M, Alvissalim M, Sanjaya F, Jatmiko W. Body gesture based control system for humanoid robot. Proceedings of the International Conference on Advanced Computer Science and Information Systems; 2012 April 16–19, Antayla, Turkey. c2012. p. 335–342.

Matsubara T, Morimoto J. Bilinear modeling of EMG signals to extract user-independent features for multiuser myoelectric interface. Proceedings of the IEEE Transactions on Biomedical Engineering; 2013;60(8):2205–2213.

Mayherry R, Jacques J. Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In: McNeill D, editor. Language and gesture. Cambridge (UK): Cambridge University Press; 2000. Chapter 10. p. 199–214.

Mitra S, Acharya T. Gesture recognition: a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 2007;37(3):311–324.

Moore D, Essa I, Hayes M. Exploiting human actions and object context for recognition tasks. Proceedings of the 7th IEEE International Conference on Computer Vision; 1999 Sep 20–27; Corfu, Greece. Piscataway (NJ): IEEE; c1999. p. 80–86.

Mortimer B, Zets G, Mort G, Shovan C. Implementing effective tactile symbology for orientation and navigation. Proceedings of the International Conference on Human Computer Interaction; 2011 July 9–14; Orlando, FL. Heidelberg (Germany): Springer-Verlag Berlin; c2011. p. 321–328.

Murthy GR, Jadon R. A review of vision based hand gesture recognition. International Journal of Information Technology and Knowledge Management. 2009 Jul–Dec 2(2):405–410.

Muto Y, Takasugi S, Yamamoto T, Miyake Y. Timing control of utterance and gesture in interaction between human and humanoid robot. The 18th IEEE International Symposium on Robot and Human Interactive Communication; 2009 Sep 27–Oct 2; Toyama, Japan.

Neely H, Belvin R, Fox J, Daily M. Multimodal interaction techniques for situational awareness and command of robotic combat entities. Proceedings of

the 2004 IEEE Aerospace Conference; 2004 Mar 6–13; Big Sky, MT. Piscataway (NJ): IEEE; c2004. p. 3297–3304.

O'Brien B, Karan C, Young S. FOCU:S—future operator control unit: soldier. In: Gehardt G, Gage D, Shoemaker C, editors. Proc. SPIE 7332, Unmanned Systems Technology XI; 2008 Apr 30; Orlando, FL. Bellingham (WA): International Society for Optics and Phonics; c 2009. p. 7332OP. doi:10.1117/12.818495.

Olufs S, Vincze M. A simple inexpensive interface for robots using the Nintendo Wii Controller. Proceedings of the 2009 International Conference on Intelligent Robots and Systems; 2009 Oct 10–15; St. Louis, MO. Piscataway (NJ): IEEE; c1999. p. 473–479.

Ong S, Ranganath S. Automatic sign language analysis: A survey and the future beyond lexical meaning. Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;27(6):873–891.

Ou S, Grupen R. From manipulation to communicative gesture. Amherst (MA): University of Massachusetts, Department of Computer Science, Laboratory for Perceptual Robotics; 2009. Technical Report.

Oz C, Leu M. Linguistic properties based on American sign language isolated word recognition with artificial neural networks using a sensory glove and motion tracker. Neurocomputing. 2007 Oct 16–18;70:2891–2901.

Park H, Kim E, Jang S, Park S, Park M, Kim H. HMM-based gesture recognition for robot control. Pattern Recognition and Image Analysis-Lecture Notes in Computer Science. 2005a;3522:607–614.

Park H, Kim E, Kim H. Robot competition using gesture based interface. Innovations in Applied Artificial Intelligence, 18th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE, LNAI 3533; 2005b; Bari, Italy. London (England): Springer Verlag; c2005b. p. 131–133.

Patel S, Abowd G. A 2-way laser-assisted selection scheme for handhelds in a physical environment. In: Dey AK, Schmidt A, McCarthy JF, editors. UbiComp 2003. Proceedings of the 5th International Conference on Ubiquitous Computing; 2003 Oct 12–15; Seattle, WA. LNCS 2864. London (England): Springer Verlag; c2003. p. 200–207.

Pavlovic V, Sharma R, Huang T. Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;19(7):677–695.

Pettitt R, Carstens C, Elliott L. Speech-based robotic control for dismounted soldiers: evaluation of visual display options. Aberdeen Proving Ground (MD): Army Research Laboratory Human Research and Engineering Directorate; 2014 May. Technical Report No.: ARL-TR-6946; ADA601525.

Perzanowski D, Adams W, Schulz A, Marsh E. Toward a seamless integration in a multi-modal interface. Washington (DC): Navy Center for Applied Research in Artificial Intelligence, NRL; 2000a. Code 5510. Report No.: ADA434973.

Perzanowski D, Brock D, Adams W, Bugajska M, Schultz A, Trafton J, Blisard S, Skubic M. Finding the FOO: a pilot study for a multimodal interface. Washington (DC): Naval Research Lab Center for Applied Research in Artificial Intelligence (US); 2003. Report No.: TR ADA434963.

Perzanowski D, Schultz A, Adams W. Integrating natural language and gesture in a robotics domain. Washington (DC): Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory (US); 1998. Code 5510. Technical Report.

Perzanowski D, Schultz A, Adams W, Bugajska M, Marsh E, Trafton G, Brock D. Communicating with teams of robots. Washington (DC): Naval Research Laboratory (US); 2002. Technical Report.

Perzanowski D, Schultz A, Adams W, Marsh E. Using a natural language and gesture interface for unmanned vehicles. Washington (DC): Navy Center for Applied Research in Artificial Intelligence, NRL; 2000b. Code 5510. Report No.: ADA435161.

Phillips E, Ososky S, Swigert B, Jentsch F. Human-animal teams as an analog for future human-robot teams. Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting. 2012;56(1):1553–1557.

Phillips E, Rivera J, Jentsch F. Developing a tactical language for future robotic teammates. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2013 Sep;57(1):1283–1287.

Pomranky-Hartnett G, Elliott L, Mortimer B, Mort G, Pettitt R, Zets G. Soldier based assessment of dual-row tactor displays during simultaneous navigation and robot monitoring tasks. Aberdeen Proving Ground (MD): Army Research Laboratory Human Research and Engineering Directorate (US); 2015 Aug. Technical Report No.: ARL-TR-7397.

Rabiner L. A tutorial on Hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989;77(2):257–286.

Rabiner L, Juang B. An introduction to hidden Markov models. IEEE ASSP Magazine. 1986;3:4–16.

Raheja J, Shyam R, Kumar U. Hand gesture capture and recognition technique for real-time video stream. In: Hamza MH, editor. ASC 2009. The 13th IASTED International Conference on Artificial Intelligence and Soft Computing; 2009 Sep 7–9, Palma de Mallorca, Spain. Calgary (Canada): ACTA Press; c2009.

Raheja J, Shyam R, Kumar U, Prasad P. Real-time robotic hand control using hand gestures. 2010 Second International Conference on Machine Learning and Computing; 2010 Feb 9–11; Bangalore, India. Piscataway (NJ): IEEE; c2010. p. 12–16.

Rao V, Mahanta C. Gesture-based robot control. Proceedings of the Fourth International Conference on Intelligent Sensing and Information Processing; 2006 Oct 15–Dec 18; Bangalore, India. Piscataway (NJ): IEEE; c2006. p. 145–148.

Rautaray SS, Agrawal A. Vison based hand gesture recognition for human computer interaction: a survey. Artificial Intelligence Review, 2012 Nov. [accessed 2014 Oct 2]. http://link.springer.com/article/10.1007/s10462-012-9356-9#page-1.

Redden E, Elliott L, Barnes M. Robots: the new team members. In: Coovert M, Foster L, editors. The psychology of workplace technology. Bowling Green (OH): Society for Industrial and Organization Psychology: Frontier Series; 2013.

Rogalla O, Ehrenmann M, Zollner R, Becher R, Dillmann R. Using gesture and speech control for commanding a robot assistant. Proceedings of the 2002 IEEE International Workshop on Robot and Human Interactive Communications; 2002 Sep 25–27; Berlin, Germany.

Ruttum M, Parikh S. Can robots recognize common marine gestures? Proceedings of the 42nd South Eastern Symposium on System Theory; 2010 Mar 7–9; University of Texas at Tyler, Texas..

Saffer D. Designing gestural interfaces. Sebastopol (CA): O'Reilley Media Inc.; 2008.

Schermerhorn P. Natural language dialogue for supervisory control of autonomous ground vehicles. Arlington (VA): Office of Naval Research, ONR Code 341, 2011. Technical Report No.: 20110519075.

Shah M, Jain R, editors. Motion-based recognition. Computational imaging and vision series. Dordrecht (Nethlands): Springer-Netherlands; 1997.

Singh R, Seth B, Desai U. A framework for real time gesture recognition for interactive mobile robots. Proceedings of International Symposium on Circuits and Systems; 2005 May; Kobe, Japan.

Skubic M, Chronis G, Matsakis P, Keller J. Generating linguistic spatial descriptions from sonar readings using the histogram of forces. In: Proceedings of the 2001 IEEE International Conference on Robotics and Automation; 2001a May 21–26; Seoul, Korea. Piscataway (NJ): IEEE;c2001. p. 485–490.

Skubic M, Chronis G, Matsakis P, Keller J. Spatial relations for tactical robot navigation. In: Gerhart GR, Shoemaker CM, editors. Proc. SPIE 4364, Unmanned Ground Vehicle Technology III; 2001b Apr 16; Orlando, FL. Bellingham (WA): International Society for Optics and Phonics; c2001. doi: 10.1117/12.439998.

Skubic M, Perzanowski D, Blisard S, Schultz A. Spatial language for human-robot dialogs. Proceedings of Systems, Man, and Cybernetics, Part C: Applications and Reviews; IEEE Transactions on. 2004;34(2):154–167.

Skubic M, Perzanowski D, Schultz A, Adams W. Using spatial language in a human-robot dialog. ICRA ′02. Proceedings of the 2002 IEEE International Conference on Robotics and Automation; 2002 May 11–15; Washington, DC. Piscataway (NJ): IEEE; c2002. p. 4143–4248.

Slade J, Bowman J. Digital flight gloves. DTIC report ADB377786; 2011.

Sofge D, Bugajska M, Adams W, Perzanowski D, Schultz A. Agent-based multimodal interface for diynamically autonomous mobile robots. Washington (DC): Naval Research Laboratory (US); Center for Applied Research in Artificial Intelligence; 2003a. ADA4344975.

Sofge D, Perzanowski D, Skubic M, Cassimatis N, Trafton G, Brock D, Bugajska M, Adams W, Schultz A. Achieving collaborative interaction with a humanoid robot. Washington (DC): Naval Research Laboratory (US); Center for Applied Research in Artificial Intelligence; 2003b. Report Accession Number: ADA434972.

Song Y, Demirdjian D, Davis R. Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database. IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG2011); Santa Barbara, CA; 2011a.

Song Y, Demirdjian D, Davis R. Multi-signal gesture recognition using temporal smoothing hidden conditional random fields. IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG2011); Santa Barbara, CA; 2011b. p. 388–393.

Song Y, Demirdjian D, Davis R. Continuous body and hand gesture recognition for natural human-computer interaction. ACM Transactions on Interactive Intelligent Systems (TiiS) – Special Issue on Affective Interaction in Natural Environments. 2012;2(1):5.

Stancil B, Hyams J, Shelly J, Babu K, Badino H, Bansai A, Huber D, Batavia P. CANINE: a robotic mine dog. Warren (MI): Army Tank Automotive Research Development and Engineering Center (US); 2013. Technical Report No.: 23494.

Starner T, Pentland A. Real-time American sign language recognition from video using hidden Markov models; 1996. AAAI Technical Report FS-96-05. [accessed 2014 Oct 2]. http://www.aaai.org/Papers/Symposia/Fall/1996/FS-96-05/FS96-05-017.pdf.

St Clair A, Atrash A, Mead R, Mataric M. Speech, gesture, and space: investigating explicit and implicit communication in multi-human multi-robot collaborations. Los Angeles (CA): University of Southern California, Department of Computer Science; 2011a Mar. ADA539427.

St Clair A, Mead R, Mataric M. Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. 2001 RO-MAN. Proceedings of the 20th IEEE International Symposium in Robot and Human Interactive Communication; 2011b Jul 31–Aug 3; Atlanta, GA. Piscataway (NJ): IEEE; c2011. p. 210–216.

Stern H, Wachs J, Edan Y. Human factors for design of hand gesture human machine interaction. Proceedings of the 2006 IEEE International Conference on Systems, Man, and Cybernetics; 2006 Oct 8–11; Taipei, Taiwan. Piscataway (NJ): IEEE; c2006. p. 4052–4056.

Stiefelhagen R, Fugen C, Gieselmann P, Holzapfel H, Nickel K, Waibel A. Natural human-robot interation using speech, head pose, and gestures. IROS 2004. Proceedings of the International Conference on Intelligent Robots and Systems; 2004 28 Sep–2 Oct; Sendai, Japan. Piscataway (NJ): IEEE. c2004. p. 2422–2427.

Stokloso A. (2015). Volkswagen goes all star wars on a Golf R interior for RAD, gesture-control CES concept. Car and Driver. 2015 Jan 5 [accessed 2015

Jan 7]. 2015. http://blog.caranddriver.com/volkswagen-goes-all-star-wars-on-a-golf-r-interior-for-rad-gesture-control-ces-concept/.

Strobel M, Illmann J, Kluge B, Marrone F. Using spatial context knowledge in gesture recognition for commanding a domestic service robot. RO-MAN ′02. Proceedings of the 11th IEEE Workshop on Robot and Human Interactive Communication; 2002 Sep 25–27; Berlin, Germany. Piscataway (NJ): IEEE; c2002. p. 468–473.

Suay H, Chernova S. Humanoid robot control using depth camera. Proceedings of the 2011 6th ACM/IEEE International Conference on Human Robot Interaction; 2011 Mar 8–11; Lausanne, Switzerland. Piscataway (NJ): IEEE; c2011. p. 401.

Sugiyama J, Miura J. A wearable visuo-inertial interface for humanoid robot control. Proceedings of the 8th ACM/IEEE International Conference on Human Robot Interaction; 2013 Mar 3–6; Tokyo, Japan. Piscataway (NJ): IEEE; c2013. p. 235–236.

Sullivan M, Hubbard E, Sullivan D, Rose B. ULTRA-Vis: Collaborative squad-level tactical augmented reality (COSTAR). Lockheed Martin Missiles and Space Co.; 2010. ADB366758. Technical Report No. AFRL-RH-WP-TR – 2011-0005.

Taylor G, Fredericksen R, Crossman J, Quist M, Theisen P. A multimodal intelligent user interface for supervisory control of unmanned platforms. Presented at Collaboration Technologies and Systems Collaborative Robots and Human Robot Interaction Workshop. 2012. Denver, CO.

Taylor G, Quist M, Lanting M, Dunham C, Theisen P, Muench P. Multi-modal interaction for robotic mules. Warren (MI): Army Tank Automotive Research Development and Engineering Center (US); 2014. Report No.: 24504. ADA595822.

Tran N, Phan H, Dinh V, Ellen J, Berg B, Lum J, Alcantara E, Bruch M, Ceruti M. Wireless data glove for gesture-based robotic control. Human-Computer Interaction Novel Interaction Methods and Techniques. Lecture Notes in Computer Science. 2009;5611:271–280.

Urban M, Bajcsy P. Fusion of voice, gesture, and human-computer interface controls for remotely operated robot. Proceedings of the 2005 7th International Conference on Information Fusion (FUSION); 2005 July 25–28; Philadelphia, PA. Piscataway (NJ): IEEE; c2005.

Urban M, Bajcsy P, Kooper R, Lementec J. Recognition of arm gestures using multiple orientation sensors: repeatability assessment. Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems; 2004 Oct 3–6; Washington, DC. Piscataway (NJ): IEEE; c2004. p. 553–558.

Varcholik P, Merlo J. Gestural communication with accelerometer-based input devices and tactile displays. In: Proceedings of the 26th Army Science Conference; 2008 Dec; Orlando, FL.

Vice J. Anthrotronix, Silver Spring, MD. Personal communication, 2015 July.

Vice J, Lathan C, Sampson J. Development of a gesture-based interface for mobile computing. In: Smith MJ, Koubek R, Salvendy G, Harris D, editors. Usability evaluation and interface design: cognitive engineering, intelligent agents, and virtual reality. Lawrence Erlbaum Ass.; 2001. p. 11–15.

Vogler C, Metaxas D. A framework for recognizing the simultaneous aspects of American sign language. Computer Vision and Image Understanding. 2001;81(3):368–384.

Waldherr S, Romero R, Thrun S. A gesture-based interface for human-robot interaction. Autonomous Robots. 2000;9:151–173.

Walter M, Friedman Y, Antone M, Teller S. Vision-based reacquisition for task-level control. In: Khatib O, Kumar V, Sukhatme G, editors. Experimental Robotics. Proceedings of the 12th International Symposium on Experimental Robotics (ISER); 2010 Dec 18–21; New Dehli, India. Heidelberg (Germany): Springer-Verlag Berlin Heidelberg; c2014. p. 493–507.

Wauchope K. Eucalyptus: Integrating natural language input with ah graphical user interface. Washington (DC): Naval Research Laboratory (US); 1994. Technical Report No.: 5510-94-9711.

Wilson A, Shafer S. XWand: UI for intelligent spaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2003 Apr 5–10; Fort Lauderdale, FL. New York (NY): ACM; c2003p. 545–552.

Wolf M, Assad A, You K, Jethani H, Vernacchia M, Fromm J, Iwashita Y. Decoding static and dynamic arm and hand gestures from the JPL BioSleeve. Proceedings of the 2013 IEEE Aerospace Conference; 2013 Mar 2-9; Big Sky, MT. Piscataway (NJ): IEEE; c2013. p. 1–9.

Wongphati M, Kanai Y, Osawa H, Imai M. Give me a hand—How users ask a robotic arm for help with gestures. Proceedings of the 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent

Systems; 2012 May 27–31; Bangkok, Thailand. Piscataway (NJ): IEEE; c2012. p. 64–68.

Wu X, Su M, Wang P. A hand-gesture-based control interface for a car-robot. 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2010 Oct 18–22; Taipei, Taiwan. Piscataway (NJ): IEEE; c2010. p. 4644–4648.

Wu Y, Huang T. Vision-based gesture recognition: A review. Lecture Notes in Artificial Intelligence. 1999;1739:103–115.

Xie C, Cheng J, Xie Q, Zhao W. Vision-based hand gesture spotting and recognition. International Conference on Information Networking and Automation (ICINA); 2010 Oct 18–19; Kunming, China. Piscataway (NJ): IEEE; c2010. p. 391–395.

Xu D, Chen Y, Lin C, Kong X, Wu X. Real-time dynamic gesture recognition system based on depth perception for robot navigation. Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics; 2012 Dec 11–14; Guangzhou, China. Piscataway (NJ): IEEE; c2012. p. 689–694.

Yan R, Tee K, Chua Y, Li H, Tang H. Gesture recognition based on localist attractor networks with application to robot control. IEEE Computational Intelligence Magazine; 2012; p. 64–74.

Yang H, Park A, Lee S. Human-robot interaction by whole body gesture spotting and recognition. Proceedings of the 18th International Conference on Pattern Recognition; 2006 Aug 20–24; Hong Kong. Los Alamitos (CA): IEEE Computer Society Press; c2006. p. 774–777.

Yanik P, Manganelli J, Merino J, Threatt A, Brooks J, Green K, Walker I. Use of kinect depth data and Growing Neural Gas for gesture-based robot control. Proceedings of the 2012 6th International Conference on Pervasive Computing Technologies for Healthcare; 2012 May 21–24; San Diego, CA.

List of Symbols, Abbreviations, and Acronyms

2-D	2-dimensional
3-D	3-dimensional
ANN	artificial neural network
ASL	American Sign Language
DARPA	Defense Advanced Research Projects Agency
DOF	degrees of freedom
DTW	dynamic time warping
EMG	electromyographic
FOV	field of view
FSM	finite state machine
GNG	Growing Neural Gas
GPS	global positioning system
HMM	hidden Markov modeling.
IMU	inertial measurement unit
MOPP	Mission-oriented Protective Posture
NN	neural network
PDA	personal digital assistant
RBF	radial basis function
TTP	tactics, techniques, and procedures
UAV	unmanned aerial vehicle
WOZ	Wizard of Oz

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIRECTOR
(PDF) US ARMY RESEARCH LAB
RDRL CIO L
IMAL HRA MAIL & RECORDS
MGMT

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

3 DIR USARL
(PDF) RDRL HRB AB
L R ELLIOTT
RDRL HRB DE
M BARNES
RDRL HRF D
S G HILL